1948

Statistical analysis of gene action in components of tomato yield

Joseph Bruce Griffing
Iowa State College

NOTE TO USERS

STATISTICAL ANALYSIS OF GENE ACTION IN COMPONENTS OF TOMATO YIELD

by

Joseph Bruce Griffing

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major Subject: Genetics

Approved:

Signature was redacted for privacy.
In Charge of Major Work

Signature was redacted for privacy.
Head of Major Department

Signature was redacted for privacy.
Dean of Graduate College

Iowa State College
1948

DEDICATION

To the memory of Dr. E. W. Lindstrom

TABLE OF CONTENTS

INTRODUCTION

To plant geneticists and breeders working with commercial crops, the methodology involved in studying complexly inherited characters is vitally important. Of particular interest is the possibility of resolving complex characters into component parts. An analytical procedure for this resolution allows a study of at least three major problems;

(1) Relationships (both phenotypic and genetic) among the various sub-traits,

(2) An estimation of the average gene properties for each of the components, and

(3) A method of selection involving various sub-characters.

An analytical scheme is important for both practical and theoretical phases of genetics. From a plant breeding aspect it allows selection pressure to be leveled at sub-characters, thus giving rise to the paradoxical situation that, theoretically, one can obtain greater genetic advance in yield of tomatoes, for example, by directing selection not at yield itself but at other characters - namely, components of yield. Powers (24) pointed out that by turning more attention to components of complex characters, a breeding program may be made more flexible. For example, similar yields may be built up in quite different patterns in two varieties. In this thesis varieties Devon and Matchless have essentially the same yield although Devon has many small, few loculed fruits, whereas Matchless has few large, many loculed fruits. In a breeding program selection may be directed at those sub-traits which will yield maximum genetic gain for

minimum amount of effort. Different traits may be emphasized in different inbred lines with the idea of combining these lines in hybrids.

From a theoretical point of view it is obvious that an eventual single gene analysis would be the ideal situation for a genetic study of a complexly inherited trait. However, the nature of the genes concerned, makes such a procedure almost impossible with present-day techniques. Even so, a breakdown of the complex into more simply inherited units is a step in this direction. Concerning this point Frankel (6, p.113) had this to say,

> The distinction between complex characters and characters with complex heredity is arbitrary and provisional, yet it may facilitate a useful approach where, by the resolution of a character into components, the latter present a clearer genetic picture than does the complex character itself . . . . Such analytical steps would tend to simplify, or even make possible a genetic analysis.

Investigations of genetic relationships among sub-traits and relative importance of each to the expression of the initial character are useful for a variety of reasons. (1) The information together with heritability estimates is used directly in discriminant function selection indices. (2) Genetic correlations are also useful in studying and interpreting gene effects of each trait. (3) Powers (24) indicated that the magnitude of genetic correlations may determine the plausibility of attempting to combine traits in a breeding program. The following question may arise in the case of a large negative genetic correlation between components: Should the emphasis of a breeding program be directed toward combining these two traits even though the large correlation exists, or should use of hybrid vigor be emphasized for simultaneous increase of both components?

The estimation of average gene properties in any quantitative character under consideration is important in determining appropriate breeding plans.

For example, Lush (18) has pointed out that whether specific combining ability is due to epistasis or over-dominance would greatly alter the breeding program which should be conducted.

Only by sound statistical techniques can reliable estimates of average gene behavior be obtained and these will lead to a clearer understanding of true general inheritance involved. Setting up gene models based on estimated gene properties allows one to predict F_1 values from those of its parents. The reliability of the prediction process depends on how well the gene models actually fit the data.

The third problem, methods of selection, is important as it implies simultaneous selection for most or all components to give maximum genetic improvement. The method used in this study for combining components into one selection index is that of discriminant functions. The concept was applied to plant breeding by Smith (27) in 1936, but since then no other real use has been made of this technique in plant breeding although it has considerable theoretical value and is apparently being used in animal breeding with some success. (See Hazel (12) and Lush (18)) Probably the main reason has been the apparently difficult and tedious task of estimating genetic correlations and heritability. It is hoped that the presentation found in this thesis may aid in this respect.

It is obvious that no one experiment can properly evaluate all of these problems. To do this a series of experiments with specially chosen material for each phase should be conducted over several seasons. This thesis is primarily concerned with the problem of estimating average gene action in characters which are inherited in a multigenic manner. Therefore choice of material and emphasis on presentation and development of techniques are

directed toward this problem. However the methodology used herein for gene estimation allows the possibility of investigating the two closely related topics mentioned above and a brief discussion of these is included.

STATISTICAL TECHNIQUES

Estimation of Average Gene Action in Quantitative Characters

Characters studied in this thesis appear, and are assumed, to be typical quantitative characters. This means that they are undoubtedly influenced by many genes. The contribution of an individual gene to the total genotypic expression is assumed small, on the average, and actually individually immeasurable with present day genetic techniques. This is so, primarily because the phenotypic expression of a quantitative trait is typically easily altered by environmental effects, particularly in the case of those complex characters whose expression depends on growth functions such as size.

Since single gene analysis is practically impossible under present experimental conditions, statistical approaches are used to study the action of these polygenes "en masse". This results in inferences about the average properties of the genes concerned.

The method of estimating average gene action in this study is to fit the experimental data to various types of gene models and to choose the model which fits the data best. Various statistical criteria are used to aid in determining; (1) whether average gene action is arithmetically or logarithmically cumulative; (2) the average inter-loci interaction

(epistatic) effect; and (3) the average direction and magnitude of the dominance effect.

In reducing the total gene action involved in the expression of a complex character to a simple gene model, it cannot be assumed that any particular gene actually behaves in the manner prescribed by the chosen model. It is supposed that the quantitative genes, variously known as multiple factors, polygenes, minor genes and in some instances as modifiers, have essentially the same diversified action exemplified by genes of larger effect. It has been suggested that these polygenes differ in action from qualitative or oligogenes only in magnitude of effect. However it is also assumed, and has been demonstrated, that groups of genes (polygenes) when handled en masse have average properties which are consistent and measurable. Estimation of these group parameters is the objective of a statistical analysis and these parameters indicate the properties of the associated gene model.

Most of the statistical methods proposed in the past have involved a description of frequency distributions resulting from segregating populations (F_2, B_1, B_2, F_3 progenies, etc.) by means of the first, second, and third moments. Probably the most popular and valuable method so far devised involving variances and covariances was first outlined by Fisher, et al (5) in 1932, and further amplified by Panse (21). Related phases were extended to F_3's and F_4's by Khambanonda (15). Because of the relative importance of the variance-covariance method it will be described briefly before describing the techniques used in this thesis which are based largely on non-segregating generations. After the two methods have been discussed, a comparison will be made on theoretical and practical grounds.

Variance - covariance method

The variance - covariance techniques, as applied to segregating populations, are aimed essentially at an indirect method of separating the total variance among segregating individuals, into component parts for the specific purpose of estimating the degree of dominance (ignoring epistatic and linkage effects).

The first step involves identifying and separating genotypic and environmental components. (Genotypic variance is that generated by differences in genotypes among the various individuals). Next, the genotypic variance is fractionated into additive and non-additive portions, and the non-additive component is taken to estimate the amount of dominance. To make these separations one must consider the following populations.

Let: Parents be denoted as,

P_{1i} $\quad i = 1, 2, \cdots, k$ plants,

P_{2i} $\quad i = 1, 2, \cdots, k$ plants,

F_1 population be denoted as,

F_{1i} $\quad i = 1, 2, \cdots, k$ plants,

F_2 population be denoted as,

F_{2i} $\quad i = 1, 2, \cdots, k$ plants.

If an F_2 plant is selfed or crossed with a plant of the same genotype, the progenies are designated as F_3's. For example, the F_3 of selfed F_{21} may be denoted as, $_iF_{(21,21)}$ $\quad i = 1, 2, \cdots, m$; where there are "m" plants in the $F_{(21,21)}$ family.

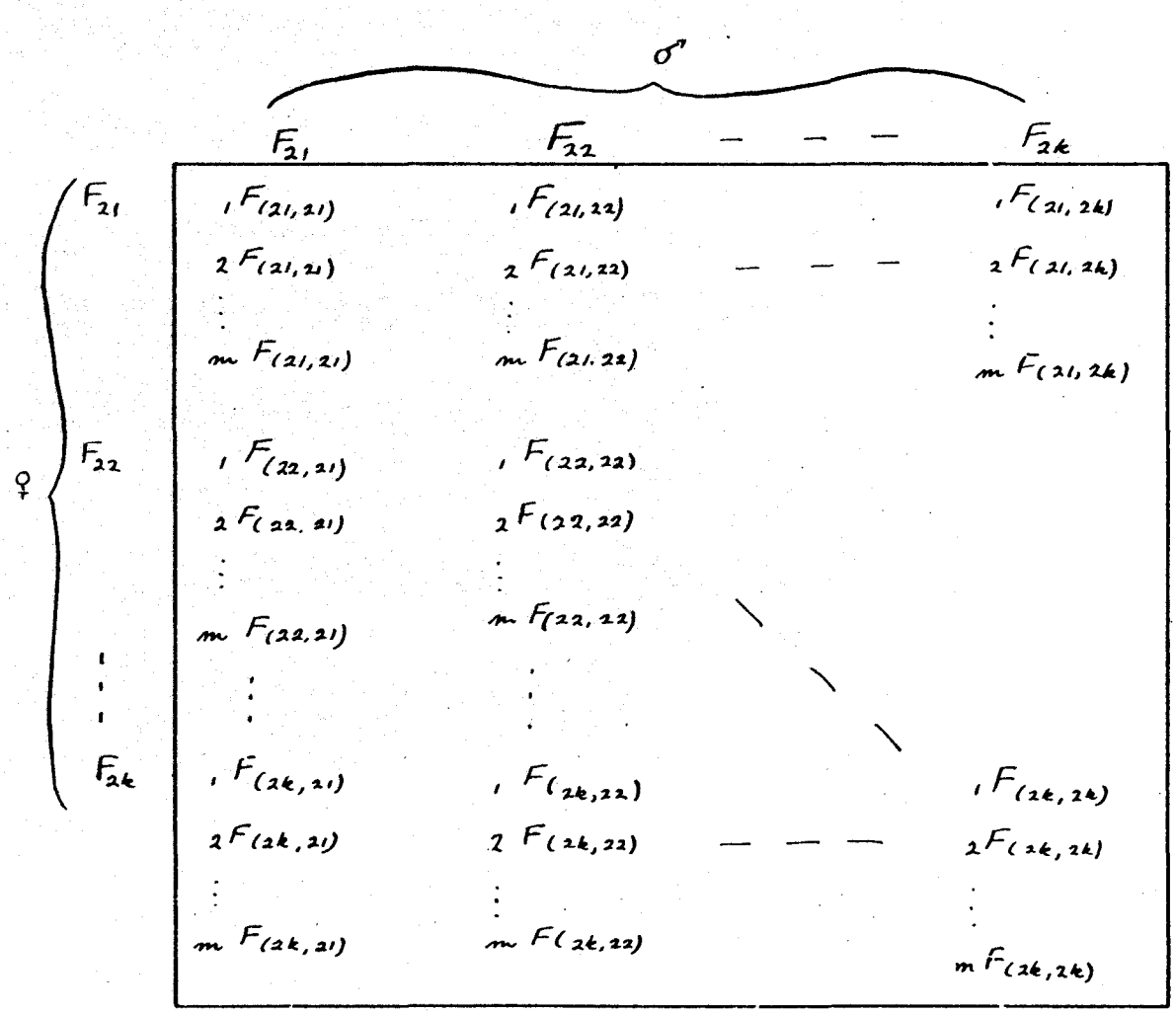
Biparental progenies result from crossing two F_2 plants chosen at random. For example, the biparental progeny (BP) resulting from crossing F_{21} with F_{22} may be designated as $_iF_{(21,22)}$ $\quad i = 1, 2, \cdots, m$; where there are

"m" biparental progenies.

These third generation populations may be more easily visualized by considering Table 1 which illustrates progenies resulting from all possible matings of F_2 individuals.

Table 1

F_3 and biparental progenies resulting from considering all possible F_2 matings



F_3 families are the families on the main diagonal to the right, which result from the mating of two like F_2's. In considering biparental progenies one must take into account all families in Table 1.

The above notation is used only to designate clearly the various populations which must be considered. For discussion of expected values a simpler form of notation will be used. It is necessary to consider the expected values for means, variances and covariances.

a. **Means:**

1. For a genetically non-segregating population consider the P_1.

Let: X_i = observed phenotype of the i^{th} plant

G_1 = genotypic component of the i^{th} plant

E_i = non-genotypic or environmental component of the plant, where E_i is a random variable with mean 0 and $E(E_i G_1) = 0$.

Then, $X_i = G_1 + E_i$

The expected value of the mean of k plants over possible environments is as follows:

$$E\{\bar{X}\} = E\left\{\frac{1}{k}\sum_{i}^{k} X_i\right\} = \frac{1}{k}\sum_{i}^{k} E(G_1 + E_i)$$

Since $E\{G_1\} = G_1$ and $E\{E_i\} = 0$

then $E(\bar{X}) = G_1$ where the gene pairs have contributions as shown on page 11.

In a similar manner it may be shown that the expected value for a phenotypic mean of any non-segregating population is the genotypic constant associated with that line or generation.

2. For an example of a genetically segregating population consider the F_2.

Let: $X_i = G_i + E_i$ where X_i is the phenotypic value of the i^{th} F_2 plant.

Then $E\{\bar{X}\} = \frac{1}{k}\sum_i^k G_i = \bar{G}$, since $E\{X_i\} = G_i$ and $E\{E_i\} = 0$.

In like manner the expected value of F_3 and biparental progeny means are equal to their respective average genotypic values.

b. Variances.

1. For a genetically non-segregating population consider P_1.

Let: $X_i = G_1 + E_i$ where X_i represents the phenotypic value of the i^{th} plant.

Then $\{X_i - \bar{X}\} = \{(G_1 - G_1) + (E_i - \bar{E})\}$,

So that $E\left\{\sum_i^k (X_i - \bar{X})^2\right\} = E\left\{\sum_i^k (E_i - \bar{E})^2\right\}$.

Since $E\left\{\sum_i^k (E_i - \bar{E})^2\right\} = (n-1)\sigma_E^2$

then $(n-1)\,V_{(X)} = (n-1)\,\sigma_E^2$

and $V_{(X)} = \sigma_E^2$

Thus the expected value of the variance of any non-segregating population is simply an environmental variance.

2. For a genetically segregating population consider an F_2 population.

From 1.b., $\{X_i - \bar{X}\} = \{(G_i - \bar{G}) + (E_i - \bar{E})\}$

So that $E\left\{\sum_i^k (X_i - \bar{X})^2\right\} = E\left\{\sum_i^k [(G_i - \bar{G}) + (E_i - \bar{E})]^2\right\}$

Since $E\{(G_i - \bar{G})(E_i - \bar{E})\} = 0$,

then $E\{\sum_i^k (X_i - \bar{X})^2\} = \sum_i^k (G_i - \bar{G})^2 + \sum_i^k (E_i - \bar{E})^2$,

and $V_{(X)} = \sigma_G^2 + \sigma_E^2$

Thus the expected value of the variance of a genetically segregating population has two components, the genotypic and environmental.

c. Covariances.

1. Consider the covariance of F_2 parental means and means of F_3 progeny. (CV_{F_3})

Let: Y_i = mean of F_3 progeny from i^{th} F_2 parent (which itself has phenotypic value X_i)

Since $E\{Y_i\} = E\{X_i\} = G_i$,

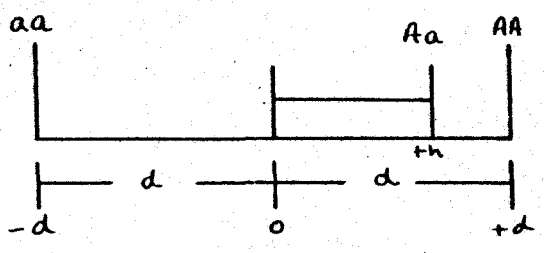
then $E\{\sum_i^m (X_i - \bar{X})(Y_i - \bar{Y})\} = \sum_i^m (G_i - \bar{G})(G_i - \bar{G}) = (n-1)\sigma_G^2$

and $CV_{F_3} = \sigma_G^2$

In an analogous manner it can be demonstrated that the expected covariance for F_2 parental mean values and means of biparental progenies (CV_{BP}) is equal to a genotypic variance or covariance free from an environmental component. This fact, that the covariances do not contain the random error effect, is important in that it allows the separation of genotypic and environmental variances.

To illustrate the separation, one must yet consider a one gene model which will give algebraic values for variance and covariance terms.

Consider the following gene model, used by Fisher et al (5):



Let: AA = +d
Aa = h
aa = -d

The various gene phases are measured from the midpoint between the two homozygous forms.

The following Table 2 illustrates the matings and their corresponding frequencies which are necessary in calculating algebraic values for variances and covariances. Those circled are the crosses yielding F_3 progeny. All crosses are used in determining CV_{GP} .

Table 2

All possible matings, and their frequencies involving monohybrid F_2 individuals.

♀ \ ♂	+d, 1AA	h, 2Aa	-d, 1aa
+d, 1AA	1[AA × AA]	2[AA × Aa]	1[AA × aa]
h, 2Aa	2[Aa × AA]	4[Aa × Aa]	2[Aa × aa]
-d, 1aa	1[aa × AA]	2[aa × Aa]	1[aa × aa]

Using algebraic values, true genotypic variances and covariances for a single gene model are as follows:

$$V_{F_2} = \frac{1}{4}(2d^2+h^2) \qquad CV_{F_3} = \frac{1}{4}(2d^2+\frac{h^2}{2}) \qquad CV_{BP} = \frac{1}{4}(d^2)$$

The next step is to fit the information at hand into a technique for separating variance components.

In populations in which genetic segregation occurs, family structure must be obtained, i.e., at least carried to F_3's, before the genotypic and environmental portions can justifiably be separated. The procedure suggested by Fisher et al (5) involves the variance and two covariance terms as follows:

Since $V_{F_2} = \sigma_G^2 + \sigma_E^2$ where $\sigma_G^2 = \frac{1}{4}(2d^2+h^2)$ with a single gene model,

then $V_{F_2} = \frac{1}{4}(2d^2+h^2) + \sigma_E^2$

Also $CV_{F_3} = \frac{1}{4}(2d^2+\frac{h^2}{2})$,and $CV_{BP} = \frac{1}{4}(d^2)$

By taking twice the difference between the covariance terms it is possible to estimate the genotypic component of the F_2 as follows:

$$2\left\{CV_{F_3} - CV_{BP}\right\} = 2\left\{\frac{1}{4}(2d^2+\frac{h^2}{2}) - \frac{1}{4}(d^2)\right\} = \frac{1}{4}(2d^2+h^2)$$

The next step is to separate the genotypic variance, $\frac{1}{4}(2d^2+h^2)$, into the additive ($\frac{d^2}{2}$), and non-additive, ($\frac{h^2}{4}$), portions. In the single gene case the non-additive fraction is a result of dominance deviations only.

This can be done by considering both parents of the CV_{BP}. The value $CV_{BP} = \frac{d^2}{4}$ is obtained by considering the covariance of mean bi-parental progeny on one parent. When covariances for both male and female

parents are added, then the result is an estimate of the additive portion of the genotypic variance, i.e., $CV_{BP}\,\sigma^2 + CV_{BP}\,\varphi = \frac{d^2}{4} + \frac{d^2}{4} = \frac{d^2}{2}$ The remaining genotypic variance is the non-additive portion and may be considered as an estimate of the dominance effect.

There are other methods of estimating some of these components, and they are as follows:

(1) Charles and Smith (1) suggest that an estimation of environmental variance may be obtained from the non-segregating populations P_1, P_2, and F_1. Then assuming that a correlation exists between means and variances an appropriate value can be estimated for F_2 environmental variance. Panse (21) points out the dangers involved in this rough approximation.

(2) Panse (21) states, "The genetic (our terminology is additive) portion of the F_2 variance can be estimated from the regression of F_3 progenies (mean) on F_2 parents." This statement has lead to some confusion because actually the regression of mean F_3 progenies on parental F_2 values is, $b = \frac{2d^2 + \frac{1}{2}h^2}{2d^2 + h^2 + 4\sigma_e^2}$ and the best interpretation that could be made of this quantity is that it is an underestimate of the genotypic variance. Possibly what he meant was that the genetic portion of F_2 variance could be estimated from the regression of mean biparental progenies from F_2 parents on the biparental F_2 parents. This would give the genetic or additive variance if the procedure outlined previously is followed. However, F_3 progenies, and progenies resulting from random mating of F_2 individuals (biparental progenies) form entirely two different populations.

In a similar manner it can be shown with the one gene model, with F and F progenies all continuous from single F plants, and assuming original

parent homozygous that the following values exist.

1. Regression of F_2 parents on F_3 progeny means = $\frac{2d^2 + h^2/2}{2d^2 + h^2/4 + 4\sigma_E^2}$

2. Regression of mean of F_3 progenies on mean of F_4's = $\frac{2d^2 + h^2/8}{2d^2 + h^2/16 + 4\sigma_E^2}$

3. Regression of mean of F_4 progenies on mean of F_3's = $\frac{2d^2 + h^2/8}{2d^2 + h^2/4 + 4\sigma_E^2}$

It would appear to the writer that these regressions do not have a straight forward interpretation, notwithstanding their relatively frequent use.

(3) Another method of measuring dominance was proposed by Fisher, et al (5) in 1932. It consisted of measuring the skewness of a distribution resulting from a segregating population by the third moment about the mean. The assumption was made that dominance was associated with an asymmetrical curve. However, skewness can be caused by a host of other factors such as logarithmic action of genes, metrical bias, an approach to a physiological limit of character expression, and epistasis. Also, since the deviations from the mean are cubed, one very erratic observation may largely determine the direction and magnitude of the third moment.

A discussion of the difficulties encountered in using segregating populations with the variance-covariance method from both practical and theoretical points of view will occur at the end of this section. Also a comparison will be made of this method with the constant parent regression scheme used in this thesis.

Constant parent regression method

The constant parent regression method as a basis for estimating gene

action in this study is based entirely on non-segregating populations - P_1, P_2, and F_1 generations in this study. The scheme is perfectly valid as each of the three gene phases enter into the estimation process through one of the generations. It is only necessary to develop the required theoretical techniques, subject them to test with experimental material, and then compare the results with those obtained by other methods, as to reliability of information acquired and practicability and ease of obtaining this information.

The basic concept of using constant parent regressions and second order regressions was first ingeniously elaborated by Hull (14) and used in discussions of overdominance in connection with heterosis. Further ramifications such as construction of gene models illustrating various types of gene interactions, and use of regression analysis of variance components were presented by Griffing (8). This section will briefly review the methodology so far available for P_1 - F_1 data and then present further ramifications and tests.

Necessary material for the constant parent regression method include a set of varieties or lines that are relatively homozygous, i.e., inbred lines. These parents must be mutually inter-fertile so that all possible F_1's are available as illustrated in the following Table 3.

Notation:

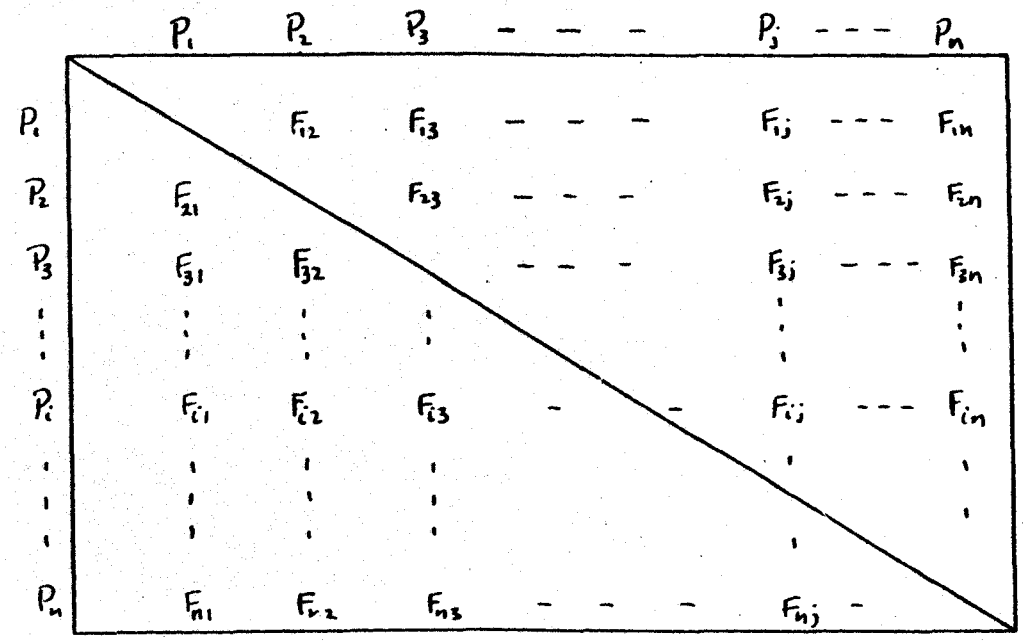
P_i = i^{th} parent

F_{ij} = F_1 (or hybrid) of i^{th} and j^{th} parents

F_{ij} = F_{ji}

Table 3

Combination of n inbred lines to give all possible F_1's



Points:

(1) "n" parents are crossed in all possible combinations to give $\binom{n}{2} = \frac{n!}{2!(n-2)!}$ hybrids.

(2) Since there are "n" parents there are "n" constant parent groups, for example:

(a) The i^{th} row would consist of the i^{th} constant parent group and would have,

1. P_i as constant parent,
2. n-1 hybrids (F_{i1}, F_{i2}, ... , F_{in}) i.e., F_1's resulting from P_i crossed with all other parents.

With this constant parent scheme as a basis, gene models are constructed

involving various intra-locus and inter-loci relationships. For simplicity models involving two gene pairs are used, as all interactions of interest may be demonstrated with two sets of genes. Such interactions will include, no dominance, complete dominance and over-dominance, for both genes acting in a plus direction and also when one gene acts in positive manner and the other negatively. Epistasis is then considered in both plus and minus conditions and in relation to dominance. All models assume effects of genes are combined in an arithmetically cumulative manner. Logarithmic gene action will be considered later. With models of two gene pairs, four parents are possible. This provides four constant parent groups each having three F_1's as follows: (See Table 4). For purposes of this investigation we shall only consider those results which folow from two gene pair models.

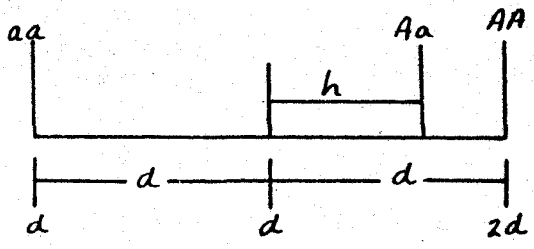
Table 4

General arrangement of P_1's and all possible F_1's for two gene models

	P_1 aabb	P_2 AAbb	P_3 aaBB	P_4 AABB
P_1 aabb		Aabb	aaBb	AaBb
P_2 AAbb	Aabb		AaBb	AABb
P_3 aaBB	aaBb	AaBb		AaBB
P_4 AABB	AaBb	AABb	AaBB	

The algebraic notation is a modification of that of Fisher, et al (5)

and is illustrated as follows:



$aa = 0 = bb$
$Aa = d(1+h) = Bb$
$AA = 2d = BB$

Dominance deviation is "hd" so that "h" may be designated as follows:

$h = 0$ no dominance

$-1 < h < 0$ negative dominance

$0 < h < +1$ incomplete positive dominance

$h > +1$ and $h < -1$ super- or over-dominance

$P_1 - F_1$ table with algebraic values is as follows Table 5: (no epistasis)

Table 5

Algebraic values for two gene model of P_1's and F_1's. Dominance considered, without epistasis. Both genes acting in positive manner.

	P_1 0	P_2 2d	P_3 2d	P_4 4d
P_1 0		$d(1+h)$	$d(1+h)$	$2d(1+h)$
P_2 2d	$d(1+h)$		$2d(1+h)$	$d(3+h)$
P_3 2d	$d(1+h)$	$2d(1+h)$		$d(3+h)$
P_4 4d	$2d(1+h)$	$d(3+h)$	$d(3+h)$	

Two main statistical tests are used to estimate polygene properties (i.e., dominance effect in the first series - without epistasis) and these are:

(1) The first test consists of a consideration of constant parent regression trends. A regression of F_1's on the variable parents is calculated for each constant parent group. For example - consider constant parent P_i.

Dependent variable	Independent variable
F_{i1}	P_1
F_{i2}	P_2
⋮	⋮
F_{in}	P_n

Thus for each parent (P_i) a c.p.r. (constant parent regression) is calculated. As will be shown later, the trend (increasing, decreasing, or no trend) of c.p.r. coefficients relative to the value of the parents will give information regarding dominance direction and magnitude. This trend is measured by a "second order" regression, i.e., the regression of c.p.r.'s on the constant parent values.

(2) The second test involves a consideration of the position of the F_1 values in relation to the regression line. Both the variance attributable to regression and the deviations from regression mean square have genetic interpretation and may be estimated by use of regression analysis of variance components.

The principle used is that the additive portion of the genotypic variance is closely estimated (in most cases) by the portion of the average total variance among the F_1's which is attributable to the regression $\frac{(\Sigma xy)^2}{\Sigma x^2}$, and the non-additive genotypic variance is estimated by

deviations from the regression $\left\{ \Sigma d_{y \cdot x}^2 \right\}$ *. With the gene models these estimates are expected values. However, with actual data there are error variances associated with the means, so that the additive and non-additive components of genotypic variance will be estimated by components of the regression analysis of variance.

Table 6 gives the regression analysis of variance components, and, then by using the algebraic values in Table 5 for constant parent two or three groups, it is possible to demonstrate the separation as shown in Table 7.

Table 6

Explanatory regression analysis of variance used for obtaining components.*

Source	dfs	Components
Regression	1	E + B
Deviations from regression	n-3*	E + D
Error		E

* n is total number of parents

$E = \hat{\sigma}^2_{\bar{F}_1}$

$D = \left\{ \frac{\Sigma d_{y \cdot x}^2}{n-2} \right\} - \hat{\sigma}^2_{\bar{F}_1}$

$B = \left\{ \frac{(\Sigma xy)^2}{\Sigma x^2} \right\} - \hat{\sigma}^2_{\bar{F}_1}$

* Notation as used by Snedecor (28)

The question at issue is to determine the relative importance of the three components of the total variance among the F_t's within a group. The three components are the additive, non-additive, and environmental variances and we are concerned primarily with the relative magnitudes of these components. A good idea of these relative magnitudes is obtained by considering E, D and B as percentages of E + D + B.

It is possible, then, to separate dominance from the additive effects for constant parent groups two and three (both genes acting in a plus direction). However, in some cases the variation attributable to regression includes a portion of the non-additive variation (i.e., constant parent groups one and four), and this is one of the main reasons for working out the gene models - to see actually what does happen to c.p.r. trends, and to the components of regression analysis of variance, with various gene relationships.

Thus two different sets of criteria, trend and components, are provided, both of which are functions of the genic interactions involved and the testing of hypotheses is strengthened.

Both genes acting in one direction; dominance considered without epistasis.

Table 7 gives the general algebraic solutions for both genes acting in one direction. The column headed "A" is the portion of total variation among F_t's removed by regression. P_t-F_t table for this model has already been given. (Table 5)

Table 7

General algebraic values for various statistics used in determining gene properties. Both genes acting in one direction.

Constant parent	Regression coefficient	$\frac{(\Sigma xy)^2}{\Sigma x^2}$	A*
$P_1 = 0$	$\frac{1+h}{2}$	$\frac{2}{3}d^2(1+h)^2$	$\frac{\frac{2}{3}d^2(1+h)^2}{\frac{2}{3}d^2(1+h)^2} = 100\%$
$P_2 = 2d$	$\frac{4d^2}{8d^2} = \frac{1}{2}$	$2d^2$	$\frac{2d^2}{2d^2+\frac{2}{3}(dh)^2}$
$P_3 = 2d$	$\frac{4d^2}{8d^2} = \frac{1}{2}$	$2d^2$	$\frac{2d^2}{2d^2+\frac{2}{3}(dh)^2}$
$P_4 = 4d$	$\frac{1-h}{2}$	$\frac{2}{3}d^2(1-h)$	$\frac{\frac{2}{3}d^2(1-h)^2}{\frac{2}{3}d^2(1-h)^2}$

* A is fraction of total variation among F_1's removed by regression.

$b_2 = -\frac{h}{4d}$ (second order regression coefficient of c.p.r. on c.p. value)

With this basic Table 7 one can substitute various values for "h" to see how the tests estimate dominance effect. Results are found in Table 8.

Table 8

Constant parent regression coefficients with various values of "h". Both genes acting in one direction.

	$h=0$	$h=+1$	$h>+1$ or $h=1+a, a>0$	$h=-1$	$h<-1$ or $h=a-1, a<0$
P = 0	.5	1.0	$1+\frac{a}{2}$	0	$\frac{-a}{2}$
P = 2d	.5	.5	.5	.5	.5
P = 2d	.5	.5	.5	.5	.5
P = 4d	.5	0	$-\frac{a}{2}$	1.0	$1+\frac{a}{2}$

Case 1: $h = 0$ (no dominance)

Regression coefficients all reduce to .5 and the regression accounts for all of the variation among the F_1's, i.e., every F_1 would fall exactly on the regression line as it connected all of the midparental values. If dominance values are increased the c.p.r. trends diverge more and more from the strictly additive (no dominance) model and more and more variance is found in the deviations from regression.

Case 2: $h = +1$ (complete positive dominance)

As the parental values increase, the c.p.r. values decrease, yielding a negative second order regression coefficient. This is then called a decreasing trend. Note that c.p.r. values extend from +1.0 to 0. For lesser values of "h" the trend would not be so extreme.

Case 3: $h>+1$ or $h=1+a$ where $a>0$ (overdominance)

As the dominance factor is increased over unity the c.p.r.'s range exceed +1 and 0. In this way a negative c.p.r. ($b = -\frac{a}{2}$) may be

associated with overdominance.

Case 4: h = -1 (negative dominance)

The important point here is that the trend is reversed. With negative dominance the trend is increasing. When $h < -1$ then just the reverse trend of $h > +1$ is found.

To sum up the procedure for estimating dominance effects when epistasis may be disregarded, the following points are listed:

1. Examine the components of regression analysis of variance to see what percentage of variation is due to deviations from regression. Also examine the mean squares to see if deviation from regression mean squares are significantly different from error.
2. Examine the c.p.r. coefficients. Direction of trend indicates plus or minus dominance. Severity of trend measures the magnitude of dominance effect.

Gene pairs acting in opposite directions; dominance considered without epistasis.

The following gene values are given for these condition :

aa = 0	bb = 2d	P_1 = aaBB = 0
Aa = d(1+h)	Bb = d(1-h)	P_2 = aabb = 2d
AA = 2d	BB = 0	P_3 = AABB = 2d
		P_4 = AAbb = 4d

Table 9 of algebraic values for gene estimation are as follows.

Table 9

General algebraic values for various statistics used in determining gene properties. Genes acting in opposite and balanced condition.

Constant parent	Regression coefficient	$\frac{(\Sigma xy)^2}{\Sigma x^2}$	A*
$P_1 = 0$	$\frac{4d^2}{8d^2} = \frac{1}{2}$	$\frac{2}{3}d^2$	$\frac{\frac{2}{3}d^2}{\frac{2}{3}d^2 + 2(dh)^2}$
$P_2 = 2d$	$\frac{1+h}{2}$	$2d^2(1+h)^2$	$\frac{2d^2(1+h)^2}{2d^2(1+h)^2} = 100\%$
$P_3 = 2d$	$\frac{1-h}{2}$	$2d^2(1-h)^2$	$\frac{2d^2(1-h)^2}{2d^2(1-h)^2} = 100\%$
$P_4 = 4d$	$\frac{4d^2}{8d^2} = \frac{1}{2}$	$\frac{2}{3}d^2$	$\frac{\frac{2}{3}d^2}{\frac{2}{3}d^2 + 2(dh)^2}$

* A is fraction of total variation among F_1's removed by regression.

$b_2 = 0$ (for all degrees of dominance)

With this basic Table 9, various values for "h" can be substituted in the formulae for c.p.r. coefficients giving the values found in Table 10.

With no dominance, all c.p.r. coefficients reduce to .5 as before and the regressions account for all of the variation among the F_1's. As "h" takes on increasing values, the c.p.r. deviate more widely from the no-dominance scheme. The trend is erratic with $b_2 = 0$ for all values of "h". The same divergence from additive scheme is true for the component "deviations from regression". As in the previous case negative "h" values reverse the c.p.r. values which are different from .5.

Table 10

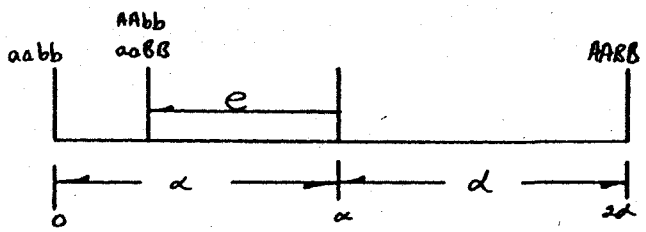
Constant parent regression coefficients with various values for "h". Genes in balanced plus and minus condition.

	h = 0	h = +1	h > 1	h = -1	h<-1
P = 0	.5	.5	.5	.5	.5
P = 2d	.5	1.0	$1+\frac{a}{2}$	.0	$-\frac{a}{2}$
P = 2d	.5	.0	$-\frac{a}{2}$	1.0	$1+\frac{a}{2}$
P = 4d	.5	.5	.5	.5	.5

For more detailed discussion of these two general types of models see Griffing (8).

General epistatic models; both dominance and epistasis considered. It is possible to set up an epistatic gene model by considering epistasis as an interaction between loci analogous to dominance as an interaction between alleles at the same locus. Such an analogy was suggested by Rasmusson (25) in presenting his interaction hypothesis.

One may construct, then, a gene model involving both epistasis and dominance in the following manner:



$-1 < e < 0$

This gives the 4 parental values: (However e can take either negative or positive values)

aabb = 0

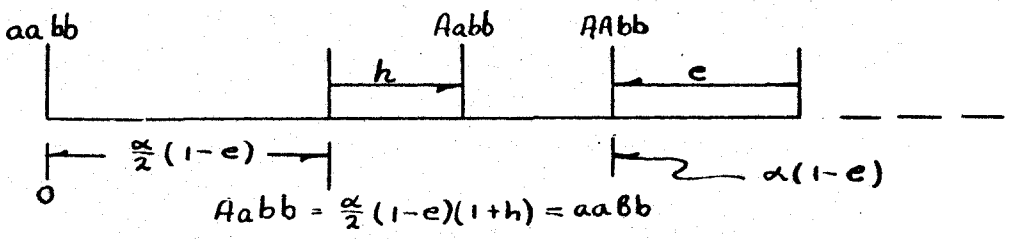
AAbb = $\alpha - \alpha e = \alpha(1-e)$

aaBB = $\alpha - \alpha e = \alpha(1-e)$

AABB = 2α

To obtain Aabb (or aaBb)

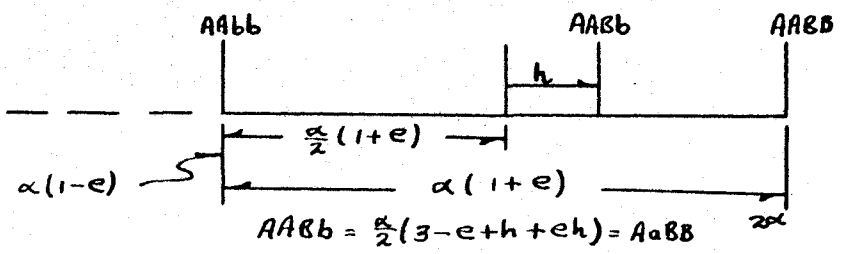
Consider: segregation of Aa on background of bb:



$Aabb = \frac{\alpha}{2}(1-e)(1+h) = aaBb$

To obtain AABb (or AaBB)

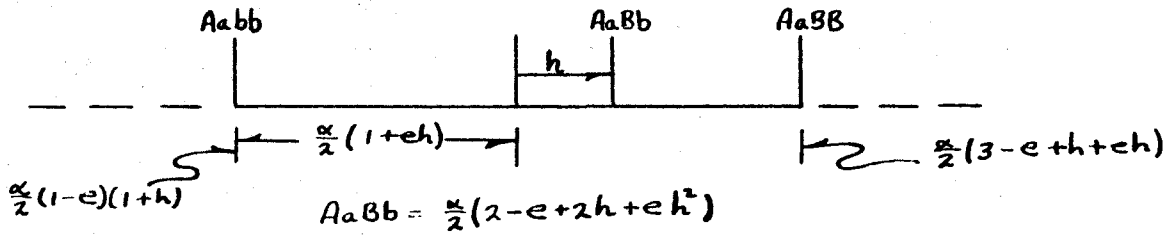
Consider: segregation of Bb on background of AA:



$AABb = \frac{\alpha}{2}(3-e+h+eh) = AaBB$

To obtain AaBb

Consider: segregation of Bb on background of Aa:



$AaBb = \frac{\alpha}{2}(2-e+2h+eh^2)$

From these values a $P_1 - F_1$ Table 11 can be constructed involving both dominance and epistasis.

Table 11

General $P_1 - F_1$ table involving both dominance and epistasis.

	P_1 aabb 0	P_2 AAbb $\alpha(1-e)$	P_3 aaBB $\alpha(1-e)$	P_4 AABB 2α
$P_1 = 0$		$\frac{\alpha}{2}(1-e+h-eh)$	$\frac{\alpha}{2}(1-e+h-eh)$	$\frac{\alpha}{2}(2-e+2h+eh^2)$
$P_2 = \alpha(1-e)$	$\frac{\alpha}{2}(1-e+h-eh)$		$\frac{\alpha}{2}(2-e+2h+eh^2)$	$\frac{\alpha}{2}(3-e+h+eh)$
$P_3 = \alpha(1-e)$	$\frac{\alpha}{2}(1-e+h-eh)$	$\frac{\alpha}{2}(2-e+2h+eh^2)$		$\frac{\alpha}{2}(3-e+h+eh)$
$P_4 = 2\alpha$	$\frac{\alpha}{2}(2-e+2h+eh^2)$	$\frac{\alpha}{2}(3-e+h+eh)$	$\frac{\alpha}{2}(3-e+h+eh)$	

The constant parent regression coefficients may be calculated for each of the constant parent groups and they are presented in Table 12.

Table 12

Algebraic values for constant parent regression coefficients for both dominance and epistasis.

Constant parent	Constant parent regression coefficient
$P_1 = 0$	$b_{P_1} = \frac{1}{2}\,\frac{\{(1+h)+e[(1+h)^2+eh(1+h)]\}}{(1+e)^2}$
$P_2 = \alpha(1-e)$	$b_{P_2} = \frac{9-2eh(2eh+3)}{6(e^2+3)}$
$P_3 = \alpha(1-e)$	$b_{P_3} = \frac{9-2eh(2eh+3)}{6(e^2+3)}$
$P_4 = 2\alpha$	$b_{P_4} = \frac{1}{2}\,\frac{\{(1-h)-e[(1-h)^2+eh(1-h)]\}}{(1-e)^2}$

It may be noted that if "e" is set equal to zero, then all algebraic values reduce to those in Table 6 (Letting $\alpha = 2d$).

As will be demonstrated in a later discussion on the estimation of dominance and epistasis variance components by a least squares solution of an F_1 prediction equation, dominance and epistatic effects cannot be separated in a direct manner when each has some value other than zero.

$$\left\{ \begin{matrix} 0 < e < 1 \\ -1 < e < 0 \end{matrix} , \quad \begin{matrix} 0 < h < +1 \\ -1 < h < 0 \end{matrix} \right\}$$

What may be done, however, and what probably yields even more accurate estimates of amounts of "e" and "h" is to make a two-way classification Table 13 of c.p.r. coefficients when various values for "e" and "h" are substituted in the constant parent regression equations. All that is necessary then is to compare the actual regression trends with those in the Table 13 and observe what values of "e" and "h" give the closest fit to the actual data.

The following characteristics of the "e" by "h" table may be noted.

(1) When $e = 0$, $h = 0$, all $b_{P_i} = .5$, as found earlier.

(2) When $e = 0$ and "h" varied (first row), the same trend occurs as found in Table 8, which is a steadily decreasing regression trend from b_{P_1} to b_{P_4}. The relative decrease depends on the degree of dominance.

(3) When $h = 0$ and "e" varied (first column), epistasis is considered alone without the confounding influence of dominance. The regression trend is quite the opposite of dominance alone, in that the regression trend steadily increases from b_{P_1} to b_{P_4}. The relative increase depends on the degree of epistasis.

(4) When h and "e" both have values, i.e., $e = .5$, $h = .5$ then the conflicting regression trends result in a decrease from b_{P_1} to b_{P_2} and an increase

from b_{P_3} to b_{P_4} .

Table 13

Two-way classification of various values of "e" and "h" for constant parent regression coefficients

Dominance Epistasis	h = 0	h = .5	h = 1.0
e = 0	b = .5	b = .75	b = 1.0
	b = .5	b = .5	b = .5
	b = .5	b = .5	b = .5
	b = .5	b = .25	b = .0
e = .5	b = .33	b = .63	b = .89
	b = .43	b = .39	b = .55
	b = .43	b = .39	b = .55
	b =1.0	b = .63	b = 0
e = 1.0	b = .25	b = .56	b = 1.0
	b = .38	b = .21	b = .46
	b = .38	b = .21	b = .46
	b = 00	b = .75	b = 0

A further consideration of Rasmusson's (25) interaction hypothesis can be made by merely assigning "e" negative values. The hypothesis may be summarized by stating that the phenotypic result of the addition of genes follows the law of diminishing returns. In other words genes added when the character expression is low would have much greater effect than when the character expression is near a physiological limit.

For examples to illustrate conditions of both plus and minus values of "e" we may consider the following simple cases.

(1) For $e = +1$; $h = +1$

Complementary gene action would fit this case. For a specific example one may consider purple and white sweet pea flower color.

(2) For $e = -1$; $h = +1$

Duplicate genes involved in shape of capsule in shepherd's purse illustrate this model. Either gene pair alone or both together cause complete phenotypic expression of this character.

These are extreme cases and obviously oversimplified when considering quantitative inheritance. When "e" takes on negative values, it tends to accentuate the dominance decreasing trend for the regression coefficients.

When "e" and "h" both become negative the regression trends for all practical purposes are reversed.

Epistasis considered without dominance. If one considers epistasis alone, some interesting facts arise. In Table 14 are given the algebraic values for parents and F_1's, and in Table 15 are listed the algebraic solutions for various statistics used.

Table 14

General algebraic values for P_i's and F_1's for condition of epistasis only.

	P_1 aabb 0	P_2 AAbb $\alpha(1-e)$	P_3 aaBB $\alpha(1-e)$	P_4 AABB $\alpha(1-e)$
$P_1 = 0$		$\frac{\alpha}{2}(1-e)$	$\frac{\alpha}{2}(1-e)$	$\frac{\alpha}{2}(2-e)$
$P_2 = \alpha(1-e)$	$\frac{\alpha}{2}(1-e)$		$\frac{\alpha}{2}(2-e)$	$\frac{\alpha}{2}(3-e)$
$P_3 = \alpha(1-e)$	$\frac{\alpha}{2}(1-e)$	$\frac{\alpha}{2}(2-e)$		$\frac{\alpha}{2}(3-e)$
$P_4 = 2\alpha$	$\frac{\alpha}{2}(2-e)$	$\frac{\alpha}{2}(3-e)$	$\frac{\alpha}{2}(3-e)$	

Table 15

General algebraic values for various statistics used in determining gene properties. Epistasis considered only.

Constant parent	Regression coefficient	$\frac{(\Sigma xy)^2}{\Sigma x^2}$	A*
$P_1 = 0$	$\frac{1}{2(1+e)}$	$\frac{\frac{1}{9}\alpha^4(1+e)^2}{\frac{6}{9}\alpha^2(1+e)^2}$	100 %
$P_2 = \alpha(1-e)$	$\frac{3}{2(3+e^2)}$	$\frac{3\alpha^2}{2(3+e^2)}$	$\frac{3\alpha^2}{2(3+e^2)} \Big/ \frac{1}{2}\alpha^2$
$P_3 = \alpha(1-e)$	$\frac{3}{2(3+e^2)}$	$\frac{3\alpha^2}{2(3+e^2)}$	$\frac{3\alpha^2}{2(3+e^2)} \Big/ \frac{1}{2}\alpha^2$
$P_4 = 2\alpha$	$\frac{1}{2(1-e)}$	$\frac{\frac{1}{9}\alpha^4(1-e)^2}{\frac{6}{9}\alpha^2(1-e)^2}$	100 %

* A is fraction of total variation among F_1's removed by regression.

$$b_2 = \frac{3e(e^2+1)}{\alpha(3+e^2)(1-e^2)(e^2+2)}$$

Curiously enough, the variance among the F_1's does not contain the epistatic factor "e", whereas the variance among the parents does.

There are two choices of procedure in estimating the epistatic variance and these are as follows:

Let $Y = F_1$ value ; $X =$ parental value

Example: Consider constant parent P_2 group:

$$\Sigma x^2 = \tfrac{2}{3}\alpha^2(3+e^2) \quad ; \quad \Sigma y^2 = \tfrac{1}{2}\alpha^2 \quad ; \quad \Sigma xy = \alpha^2$$

(1) First procedure: - break up the sum of squares, Σx^2, into additive and epistatic portions by:

$\frac{(\Sigma xy)^2}{\Sigma y^2} = 2\alpha^2$ portion attributable to regression (additive portion), and

$\Sigma d_{x \cdot y}^2 = \frac{2}{3}\alpha^2 e^2$ deviations from regression (epistatic portion).

These 2 portions add up to Σx^2.

(2) Second possible procedure is as follows:

$\frac{(\Sigma xy)^2}{\Sigma x^2} = \frac{3\alpha^2}{2(3+e^2)}$ which corresponds to the additive portion, and

$\Sigma d_{y \cdot x}^2 = \frac{\alpha^2 e^2}{2(3+e^2)}$ which corresponds to the epistatic portion.

These can be shown to divide the $\Sigma y^2 = \frac{1}{2}\alpha^2$ into portions which are equivalent to the first procedure, and since the second procedure was that used throughout the other models, it will continue to be used here. The first model involves reversing the regression which is valid is this case.

Regression trends have been described already for the joint "e" and "h" case with no dominance. Epistasis causes a steadily increasing regression trend. The isolation of epistatic components of variance is valid for the two intermediate parental groups, but for the two extremes the regression removes both additive and epistatic variances.

The second order regression coefficient turns out to be;

$$b_2 = \frac{3e(e^2+1)}{\alpha(3+e^2)(1-e^2)(e^2+2)}$$

Tests of significance for b_2 regression coefficients. Much importance is attached to the regression trend which can be measured (in certain cases) by the b_2 (second order regression coefficient). Therefore it is of some

interest to determine whether the regression is significantly different from zero.

An ordinary analysis of variance for regression of c.p.r. coefficients on parental values is not valid because the c.p.r. coefficients are all highly correlated.

The problem can be approached, however, by fitting constants to the following regression equation by least squares procedure.

Let: Y_{ij} = F_1 of parents P_i and P_j

$X_i = P_i$; $X_j = P_i$

$$Y_{ij} = m + a(X_i + X_j) + \rho(X_i^2 + X_j^2) + q(X_i X_j) + \epsilon_{ij} \qquad (1)$$

On partial differentiation with respect to each of the constants and setting, each equal to zero, four simultaneous equations are obtained.

$$m \cdot n + a\Sigma(X_i+X_j) + \rho\Sigma(X_i^2+X_j^2) + q\Sigma X_i X_j = \Sigma Y_{ij}$$

$$m \cdot \Sigma(X_i+X_j) + a\Sigma(X_i+X_j)^2 + \rho\Sigma(X_i+X_j)(X_i^2+X_j^2) + q\Sigma X_i X_j (X_i+X_j) = \Sigma Y_{ij}(X_i+X_j)$$

$$m \cdot \Sigma(X_i^2+X_j^2) + a\Sigma(X_i+X_j)(X_i^2+X_j^2) + \rho\Sigma(X_i^2+X_j^2)^2 + q\Sigma X_i X_j (X_i^2+X_j^2) = \Sigma Y_{ij}(X_i^2+X_j^2)$$

$$m \cdot \Sigma X_i X_j + a\Sigma X_i X_j (X_i+X_j) + \rho\Sigma X_i X_j (X_i^2+X_j^2) + q\Sigma X_i^2 X_j^2 = \Sigma Y_{ij}(X_i X_j)$$

Algebraic solution of this system of equations is accomplished separately for dominance considered alone and epistasis considered alone. The algebraic solution is nearly impossible for the joint consideration of dominance and epistasis, and therefore numerical examples are presented for this condition.

Case I - dominance considered alone

Algebraic solution gives rise to the following values for the various constants.

$m = 0$ $\qquad$ $\rho = 0$

$a = \frac{1+h}{2}$ $\qquad$ $q = -\frac{h}{4d^2}$

What is actually being done, of course, is fitting a regression surface to an entire F_2 dihybrid (using appropriate parents for each F_2 value); or using only the 12 entries in Table 5. Both sets yield the same result and the regression surface fits all points exactly.

The constant $q = -\frac{h}{4d^2}$ is actually the same as the second order regression coefficient.* This affords a perfectly valid test, then, for "q" as a sum of squares can be isolated for it and an F test made.

Of special interest is the fact that p = 0, in other words dominance does not contribute to the constant "p", with the above assumptions ($e = o$). The constant "a" = $\frac{1+h}{2}$ is equivalent to the P_1 regression coefficient but has no particular interest here.

Other equations and their least squares solutions are as follows:

$$Y_{ij} = m + a(X_i + X_j) + b(X_i + X_j)^2 + c(X_i - X_j)^2 + \epsilon_{ij} \qquad (2)$$

Values for constants are:

$m = o$ $\qquad b = -\frac{h}{16d^2}$

$a = \frac{d+h}{2d}$ $\qquad c = +\frac{h}{16d^2}$

$$Y_{ij} = m + aX_i + bX_j + qX_iX_j + \epsilon_{ij} \qquad (3)$$

Values for constants are:

$m = o$ $\qquad b = \frac{d+h}{2d}$

$a = \frac{d+h}{2d}$ $\qquad q = -\frac{h}{4d^2}$

---

* $b_2 = -\frac{h}{4d^2}$ if dominance deviation = h. (scheme used for working out least squares solutions)

$b_2 = -\frac{h}{4d}$ if dominance deviation = hd. (as has been used earlier)

Again in equation (3) the constant "q" is equivalent to b_2 (second order regression coefficient) so that in absence of epistasis a solution of this sort of equation will yield a test for b_2. It is needless to say that solving for the constants in equation (3) is far simpler than for the constants in the original quadratic (1).

Hull (14) has presented the solution to an equation equivalent to (3).

Case II. Epistasis considered alone

Algebraic solution for equation (1) under the restriction that h = 0 (considering epistasis alone) yields the following values for the constants.

$$m = -\frac{\alpha e}{2} \qquad p = -\frac{e}{2\alpha(1-e^2)}$$

$$a = +\frac{1+2e-e^2}{2(1-e^2)} \qquad q = 0$$

Three constants "m", "a", and "p" all contain factor "e". Of interest is the one which can be associated with b_2 of the model which considers epistasis alone.

Consider the solution for the constants in three cases in which "e" takes on different values and "h" is held equal to zero (see Table 16).

One can see that absolute values of "p" and b_2 are similar. The relationship can be examined in more detail in Table 17.

It appears that for all intents and purposes, constant "p", although not identical algebraically is closely similar (with sign changed) to b_2

and therefore may be used to estimate the b_2 effect.

Table 16

Comparison of values of constants with that of b_2 for different values of "e".

Constants	$e=0, h=0$	$e=\frac{1}{2}, h=0$	$e=1, h=0$
m	0	$\frac{\alpha}{4}$	$\frac{\alpha}{2}$
a	$\frac{1}{2}$	$\frac{1}{\epsilon}$	$+\infty$
p	0	$-\frac{1}{3\alpha}$	$-\infty$
q	0	0	0
For b_2^*	0	$+\frac{40}{117\alpha}$	$+\infty$

$^* b_2 = \frac{3e(e^2+1)}{\alpha(3+e^2)(1-e^2)(e^2+2)}$

Table 17

Comparison of values of "p" and b_2 with different values of "e" ($h = 0$).

Constants	$e = 0$	$e = .2$	$e = .4$	$e = .6$	$e = .8$	$e = 1.0$
p	0	$-\frac{.1042}{\alpha}$	$-\frac{.2381}{\alpha}$	$-\frac{.4688}{\alpha}$	$-\frac{1.1379}{\alpha}$	$-\infty$
b_2	0	$+\frac{.1048}{\alpha}$	$+\frac{.2428}{\alpha}$	$+\frac{.4824}{\alpha}$	$+\frac{1.1111}{\alpha}$	$+\infty$

If one accepts this, then there are two constants "p" and "q" which may be used to estimate the non-linear effects of dominance and epistasis respectively when each effect is considered alone.

Case III. Presence of both dominance and epistasis

Algebraic solution is virtually impossible for this condition; however, the following numerical examples in Table 18 will suffice to point out that

when there are both "e" and "h" effects, a solution for each effect would be difficult with this equation.

Table 18

Comparing solution of constants for three numerical examples when values of "e" and "h" are varied.

Case I	Case II	Case III
Dominance - no epistasis	Epistasis - no dominance	Both dominance and epistasis
$\alpha=4, e=0, h=1$	$\alpha=4, e=\frac{1}{2}, h=0$	$\alpha=4, e=\frac{1}{2}, h=1$
$m = 0$	$m = -1$	$m = -6$
$a = \frac{3}{4}$	$a = \frac{7}{6}$	$a = \frac{19}{4}$
$p = 0$	$p = -\frac{1}{12}$	$p = -\frac{3}{8}$
$q = -\frac{1}{16}$	$q = 0$	$q = -\frac{1}{2}$

Even though $p = 0$ in Case I and $q = 0$ in Case II, the third case demonstrates that there must be interaction terms involving "e" and "h" in both "p" and "q", as it can be seen that in the third case where both dominance and epistasis are considered, "q" no longer measures only - $-\frac{h}{4d^3} = -\frac{1}{16}$ (where $\alpha = 2d$), but is changed by the element of epistasis. Likewise "p" also changes when both dominance and epistatic effects appear.

The constants and the sum of squares attributable to the constants, have genetic interpretation and are of use for above-mentioned tests only when one effect is considered in the absence of the other.

Actually, however, these are the only cases for which one would want

a test for b_2. When "e" and "h" both have some value, then, the trend is generally curvilinear so that the linear regression coefficient b_2 would have little meaning.

Logarithmic gene action

Logarithmic gene action implies that the logarithms of the genotypic values would fit the cumulatively additive scheme. Thus, by transforming to logarithms, the data can be fitted to any of the models so far presented which are based on additive gene action.

For example, the assumption is often made that the increase of fruit weight is a function of the exponential expression $W = Ae^{rt}$; where W = weight of fruit, A = initial weight, r = rate of growth (cell division), and t = time. (The actual growth curve is of a logistic form, which is a function of the above-mentioned exponential. See Lush (18).

Then genes influencing the rate of cell division (denoted as r) would fall into the class having logarithmic effects. The gene model would be as follows:

$AA = e^{2d}$

$Aa = e^{d(1+h)}$

$aa = e^{0}$

2 genes

$AABB = e^{4d}$

$AaBB = e^{2d + d(1+h)}$

$AABb = e^{2d + d(1+h)}$

$AaBb = e^{2d(1+h)}$

etc.

Thus, by merely converting to logs, the data are resolved into the additive scheme so that the log data may be fitted to any arithmetical model.

Geometric gene action, where the effects of the genes are multiplied,

would be converted to the additive scheme by logarithmic transformation. This may be shown as follows:

$$\text{Genotype} = (\text{Gene } A_1)\ (\text{Gene } A_2) - - - (\text{Gene } A_n) = \prod_{i=1}^{n} (\text{Gene } A_i)$$

$$\log \text{Genotype} = \log(\text{Gene } A_1) + \text{---} + \log(\text{Gene } A_n) = \sum_{i=1}^{n} \log(\text{Gene } A_i)$$

This scheme would include the geometric model illustrated by Charles and Smith (1) where alleles contribute the same percent increment to the expression of a trait.

When characters, controlled by logarithmic or geometric gene action, are measured in arithmetic values, it is obvious that epistasis or gene interaction exists. In other words the addition of a gene has an exponential or multiplicative effect, the value of which is a function of the rest of the genotype. By working out a simple numerical example it may be readily observed that this general type of interaction is detected in a familiar manner, that of an increasing regressive trend and significant deviations from regression.

When, however, the data are transformed, the gene action itself is transformed to the additive scheme and the interaction disappears with the transformation. This type of interaction is termed metrical bias and results when variations are measured on a different scale than is actually involved in the growth and expression of the character. Probably most types of interactions between genes are considered interactions because the scale of measurement and scale of physiological activity do not coincide. This might imply that an approach to the problem of estimating gene action would be to find a transformation of the data such that the gene action is accounted for by a purely additive no-dominance scheme, and then to interpret the needed transformation.

A more mathematical approach to the reason for considering a log function as the limiting form for a general type of gene interaction (epistasis) as the number of gene pairs becomes large, is given by Cramér (2, p.219) in extending the central limit theorem of mathematical statistics.

> If our random variable is the size of some specific organ that we are observing, the actual size of this organ in a particular individual may often be regarded as the joint effect of a large number of mutually independent causes, acting in an ordered sequence during the time of growth of the individual. If these causes simply add their effects, which are assumed to be random variables, we infer by the central theorem that the sum is asymptotically normally distributed.
>
> In general it does not, however, seem plausible that the causes cooperate by simple addition. It seems more natural to suppose that each cause gives an impulse, the effect of which depends both on the strength of the impulse and on the size of the organ already attained at the instant when the impulse is working.
>
> Suppose that we have n impulses $\xi_1, \xi_2, \cdots, \xi_n$ acting in the order of their indices. These we consider as independent random variables. Denote by x_ν the size of the organ which is produced by the impulses $\xi_1, \cdots, \xi_\nu$. We may then suppose e. g. that the increase caused by the impulse $\xi_{\nu+1}$ is proportional to $\xi_{\nu+1}$ and to some function $g(x_\nu)$ of the momentary size of the organ:
>
> $$x_{\nu+1} = x_\nu + \xi_{\nu+1} g(x_\nu)$$
>
> It follows that we have
>
> $$\xi_1 + \xi_2 + \cdots + \xi_n = \sum_0^{n-1} \frac{x_{\nu+1} - x_\nu}{g(x_\nu)}$$
>
> If each impulse only gives a slight contribution to the growth of the organ, we thus have approximately
>
> $$\xi_1 + \xi_2 + \cdots + \xi_n = \int_{x_0}^{x} \frac{dt}{g(t)}$$
>
> where $x = x_n$ denotes the final size of the organ. By hypothesis $\xi_1, \cdots, \xi_n$ are independent variables and "n" may be considered as a large number. Consider, e.g. the case of $g(t) = t$. The effect of each impulse is then directly proportional to the momentary size of the organ. In this case we thus find that <u>logx</u> is normally

> distributed. - - The corresponding frequency curve — is unimodel and of positive skewness.

Thus it may be interpreted that, if a set of genes do not have simple additive properties, but exhibit some sort of general type of gene interaction in which the effect of the gene depends not only on its own activity but on the presence of other genes in the genotype, then, assuming a large number of genes, the log of the genotype is normally distributed. In other words since gene interaction approaches a log function with large n, then log transformation should reduce the gene effects to an additive scheme.

Traits which might exhibit geometric gene action would include those which vary volumetrically or are measured in a three-dimensional manner. Included also would be traits which result in compounding sub-traits in a multiplicative manner (i.e., yield as broken-down in this study). It is conceivable that the resolution into component parts could eventually lead to the consideration of the chemical effects of individual genes, where the increase or decrease of a specific gene controlled enzyme would have proportionate effects on the expression of an immediate product.

It is not surprising, then, to find logarithmic gene action rather common among those quantitative characters which are largely functions of growth processes and generally of a complex nature.

Use of means for calculating dominance effects

Besides regression trends and component analysis one can estimate the amount of dominance directly from averages of segregating and non-segregating populations. The proper procedure is to determine the type

of gene action first, whether arithmetic or logarithmic, make appropriate transformations, and then compare F_1 (or other non-parental) means with the mid-parental values to determine the dominance effect.

1. Use of F_1 means

Phenotypic dominance can be calculated from the following formula:

$$h = \frac{F_1 - M_P}{P_2 - M_P}$$

M_P = midparent

P_2 = parent with greatest expression of character

2. Use of F_2, B_1, and B_2 means

Segregating population means may also be used for estimating dominance as follows:

For F_2: $h = \frac{2(F_2 - M_P)}{P_2 - M_P} - 2$

For B_1: $h = \frac{2(B_1 - M_P)}{P_2 - M_P} - 1$ B_1 = backcross to low parent (P_1)

For B_2: $h = \frac{2(B_2 - M_P)}{P_2 - M_P} - 3$ B_2 = backcross to high parent (P_2)

"h" values are calculated on data freed from interactions (i.e. on transformed data if gene action is found to be logarithmic) and thus measure the magnitude of dominance as accurately as possible.

An "h" value so calculated is not identified with an "h" value of any single individual gene, but with an average of many genes. It is a property of the group of genes influencing the character under consideration. This average "h" is used as the average dominance deviation in the model.

General procedure for estimation of average gene action.

The following steps are listed as the general procedure used in this

thesis for estimating average gene effects.

1. The arithmetic P_1 and F_1 means are examined and various statistics calculated such as;

   (1) c.p.r., (2) b_2, (3) components of regression analysis of variance, and (4) "h" values from F_1's, F_2, B_1 and B_2.

2. If the regression trend is decreasing, then positive dominance and arithmetic gene action are assumed. Amount of dominance is estimated by severity of regression trend, relative values of regression components, and significance of mean squares of deviations from regression, and average "h" values.

3. If the regression trend is increasing then the first problem arising is to distinguish between (1) arithmetic cumulative action with negative dominance, and (2) logarithmically cumulative gene action with or without dominance.

This is done by transforming data to logs and comparing the various statistical values with the arithmetic analysis. Then if,

(a) on transforming to logs the c.p.r. trend is drastically reduced in comparison with the trend of the arithmetic c.p.r.'s., this indicates that the arithmetic trend was due to log interaction (metrical bias) which can be removed by transformation.

(b) on transforming to logs the deviations from regression are greatly reduced, this indicates that $\Sigma d_{y \cdot x}^2$ is due to metrical bias which disappears with transforming.

(c) "h" values are irregular with arithmetic data and on transforming to logs become much more uniform throughout, this indicates that

the best gene model is logarithmic.

(d) in arithmetic analysis the c.p.r.'s are all positive but increase sharply with largest valued c.p.r. considerably over +1.00, this indicates logarithmic gene action. With negative dominance, c.p.r. values should never exceed +1.00 (disregarding errors associated with c.p.r.'s) except in the case of overdominance and then at least one c.p.r. should have a negative value.

These four different criteria may be used quite effectively to differentiate logarithmic from arithmetic gene action. The best example of this procedure is the study of fruit weight.

4. When proper basic gene action is decided upon, then estimates of amount of dominance are determined, using the appropriate data (arithmetic or log), from the three statistics:

(a) b_2 - severity of regression trend,

(b) $\Sigma d_{1\cdot x}^2$ - deviations from regression,

(c) "h" tables.

Consideration of means of segregating populations

Since an F_2 and backcrosses involving two parents are included in this study a few points should be mentioned concerning the expected values of the means.

With the single gene model used previously in the variance-covariance discussion, dominance trends can be determined in succeeding selfed generations (epistasis ignored).

With positive dominance the means of succeeding generations converge

from high F_1 value on M_p (midparent) i.e., a decreasing trend.

$\bar{F}_1$	$\bar{F}_2$	$\bar{F}_3$	$\bar{F}_4$		
$M_p + h$	$M_p + \frac{h}{2}$	$M_p + \frac{h}{4}$	$M_p + \frac{h}{8}$	--	$\lim_{n \to \infty} \bar{F}_n = M_p$

With the negative dominance the means increase, converging on M_p., i.e. increasing trend.

$\bar{F}_1$	$\bar{F}_2$	$\bar{F}_3$	$\bar{F}_4$		
$M_p - h$	$M_p - \frac{h}{2}$	$M_p - \frac{h}{4}$	$M_p - \frac{h}{8}$	--	$\lim_{n \to \infty} \bar{F}_n = M_p$

Table 19 gives the means for P_1's, F_1, F_2, B_1 and B_2 assuming arithmetically cumulative gene action. Different values are assigned to "e" and "h". Table 20 considers logarithmic gene action.

Comparison of variance - covariance and constant parent regression techniques.

The variance-covariance technique as presented early in this section was developed using a one gene model. As soon as another gene is considered in the model, two fundamental problems immediately rise, and these are:

(a) Interaction between non-allelic genes - The genotypic variance may be separated into the additive and non-additive portions, but the non-additive variance is an estimate of the combined effects of dominance and epistasis. The problem, then, is to separate dominance and epistatic components, and this has not been successfully done with this approach.

(b) Linkage - with genetically segregating populations, and assuming large number of genes concerned with inheritance of any quanti-

	h = 0	h = .5	h = 1.0
e = 0	$\bar{P}_1 = 0$ $\quad \bar{P}_2 = 2\alpha$ $\bar{F}_1 = \alpha$ $\bar{B}_1 = \frac{1}{2}\alpha$ $\quad \bar{B}_2 = \frac{3}{2}\alpha$ $\bar{F}_2 = \alpha$	$\bar{P}_1 = 0$ $\quad \bar{P}_2 = 2\alpha$ $\bar{F}_1 = \frac{3}{2}\alpha$ $\bar{B}_1 = \frac{3}{4}\alpha$ $\quad \bar{B}_2 = \frac{7}{4}\alpha$ $\bar{F}_2 = \frac{5}{4}\alpha$	$\bar{P}_1 = 0$ $\quad \bar{P}_2 = 2\alpha$ $\bar{F}_2 = 2\alpha$ $\bar{B}_1 = \alpha$ $\quad \bar{B}_2 = 2\alpha$ $\bar{F}_2 = \frac{3}{2}\alpha$
e = .5	$\bar{P}_1 = 0$ $\quad \bar{P}_2 = 2\alpha$ $\bar{F}_1 = \frac{3}{4}\alpha$ $\bar{B}_1 = \frac{5}{16}\alpha$ $\quad \bar{B}_2 = \frac{21}{16}\alpha$ $\bar{F}_2 = \frac{3}{4}\alpha$	$\bar{P}_1 = 0$ $\quad \bar{P}_2 = 2\alpha$ $\bar{F}_1 = \frac{21}{16}\alpha$ $\bar{B}_1 = \frac{9}{16}\alpha$ $\quad \bar{B}_2 = \frac{105}{64}\alpha$ $\bar{F}_2 = \frac{65}{64}\alpha$	$\bar{P}_1 = 0$ $\quad \bar{P}_2 = 2\alpha$ $\bar{F}_1 = 2\alpha$ $\bar{B}_1 = \frac{3}{4}\alpha$ $\quad \bar{B}_2 = 2\alpha$ $\bar{F}_2 = \frac{21}{16}\alpha$
e = 1.0	$\bar{P}_1 = 0$ $\quad \bar{P}_2 = 2\alpha$ $\bar{F}_1 = \frac{1}{2}\alpha$ $\bar{B}_1 = \frac{1}{8}\alpha$ $\quad \bar{B}_2 = \frac{9}{8}\alpha$ $\bar{F}_2 = \frac{1}{2}\alpha$	$\bar{P}_1 = 0$ $\quad \bar{P}_2 = 2\alpha$ $\bar{F}_1 = \frac{9}{8}\alpha$ $\bar{B}_1 = \frac{9}{32}\alpha$ $\quad \bar{B}_2 = \frac{49}{32}\alpha$ $\bar{F}_2 = \frac{25}{32}\alpha$	$\bar{P}_1 = 0$ $\quad \bar{P}_2 = 2\alpha$ $\bar{F}_1 = 2\alpha$ $\bar{B}_1 = \frac{1}{2}\alpha$ $\quad \bar{B}_2 = 2\alpha$ $\bar{F}_2 = \frac{9}{8}\alpha$

NOTE:— FOR USE OF ABOVE FORMULA, (1) ADD VALUE OF P_1 TO ALL MEANS, (2) SUBTRACT $P_2 - M_P$ FOR α

TABLE 19. EXPECTED P_1's, F_1, F_2, B_1, AND B_2 MEANS FOR VARIOUS DEGREES OF DOMINANCE AND EPISTASIS WITH TWO GENE MODEL: ARITHMETICALLY CUMULATIVE ACTION ASSUMED.

h = 0	h = .5	h = 1.0
$\bar{P}_1 = 10^0$ $\quad \bar{P}_2 = 10^{2\alpha}$ $\bar{F}_1 = 10^{\alpha}$ $\bar{B}_1 = 10^{\frac{1}{2}\alpha}$ $\quad \bar{B}_2 = 10^{\frac{3}{2}\alpha}$ $\bar{F}_2 = 10^{\alpha}$	$\bar{P}_1 = 10^0$ $\quad \bar{P}_2 = 10^{2\alpha}$ $\bar{F}_1 = 10^{\frac{3}{2}\alpha}$ $\bar{B}_1 = 10^{\frac{3}{4}\alpha}$ $\quad \bar{B}_2 = 10^{\frac{7}{4}\alpha}$ $\bar{F}_2 = 10^{\frac{5}{4}\alpha}$	$\bar{P}_1 = 10^0$ $\quad \bar{P}_2 = 10^{2\alpha}$ $\bar{F}_1 = 10^{2\alpha}$ $\bar{B}_1 = 10^{\alpha}$ $\quad \bar{B}_2 = 10^{2\alpha}$ $\bar{F}_2 = 10^{\frac{3}{2}\alpha}$
LET $\alpha = 1$ $\bar{P}_1 = 1$ $\quad \bar{P}_2 = 100$ $\bar{F}_1 = 10$ $\bar{B}_1 = 3.16$ $\quad \bar{B}_2 = 31.6$ $\bar{F}_2 = 10$	LET $\alpha = 1$ $\bar{P}_1 = 1$ $\quad \bar{P}_2 = 100$ $\bar{F}_1 = 31.6$ $\bar{B}_1 = 5.62$ $\quad \bar{B}_2 = 56.2$ $\bar{F}_2 = 17.8$	LET $\alpha = 1$ $\bar{P}_1 = 1$ $\quad \bar{P}_2 = 100$ $\bar{F}_1 = 100$ $\bar{B}_1 = 10$ $\quad \bar{B}_2 = 100$ $\bar{F}_2 = 31.6$

TABLE 20. EXPECTED P_1's, F_1, F_2, B_1, AND B_2 MEANS FOR LOGARITHMIC GENE ACTION WITH BASE 10 BASED UPON TWO GENE MODEL CASE.

tative character, linkage effects are generated and contribute an unknown amount to the variance.

(c) Segregating populations require a relatively large number of plants - i.e., for proper sampling of F_2 progenies large numbers of plants are necessary, and then to progeny test each F_2 plant adequately with F_3's and biparental progenies, the numbers would become tremendous. This makes the technique almost impossible for two reasons:

(1) From the practical standpoint, the relative amount of work and expense for the amount of information obtained would be high. This technique would be practically impossible for characters such as yield with plants having indeterminate growth. Determing yield for a plant having many small fruits is time consuming, and if there were many such plants the amount of work involved would make the project almost impossible under ordinary conditions.

(2) From a statistical point of view, difficulties would arise in designing an experiment with such large numbers of plants in which estimates would be accurate and comparable. Also difficulty would be encountered by confounding genetic effects with others such as replication effect.

On the other hand, this method may be used in connection with an actual

breeding program, where the necessary information for estimation of gene action would be obtained as a byproduct. Then, of course, there may be other circumstances which would make this method the most practical to use. Segregating populations must be used to get information on such things as detection of major genes, linkage with qualitative characters, and estimation of gene number.

Use of the $P_1 - F_1$ technique evades or does not generate most of the previously mentioned difficulties. Gene interaction, of course, would exist but may be estimated and separated from dominance effects. The linkage problem does not exist with genetically non-segregating populations. Since the P_1's and F_1's are not segregating, relatively few plants are needed to estimate the genetic parameters. The actual number needed depends on the heritability of the traits involved, and accuracy may be increased to any practical desired level by merely increasing the number of plants. Variance within a line is assumed to be entirely environmental and thus, with appropriate sampling techniques an analysis of the source of these environmental variations is provided. Experimental designing presents no problems and there is no genetic confounding. (There is, of course, a genotypic - environmental interaction which may be estimated).

One of the major advantages of the $P_1 - F_1$, or constant parent regression technique is that it has properties which liken it to the concept of factorial experimentation. With the variety of different parents involved, the method provides a wider sampling of the germ-plasm available, and thus allows broader inferences to be made from the results obtained. Segregating populations, on the other hand, generally trace back to only two parents. Information akin to interactions is available in the P_1

method since each parent is crossed onto essentially the same group of parents. This gives rise to information of individual interactions (specific combing ability) and even to group interactions such as the comparison of related lines and unrelated lines occuring in the same experiment (see Griffing (8)).

By using a number of parents, the experimenter should be able to choose those lines which, collectively, would give desired ranges of expression in all of the characters of interest. This might be difficult to do when only two parents must be chosen. The final advantage, and one of the most important points, is that the $P_1 - F_1$ methodology yields estimates directly and easily on heritability, genetic correlations and other information which can be used in determining relationships among characters and developing selection indices if desired.

Like all techniques the $P_1 - F_1$ test has shortcomings and limitations. Linkage studies, obviously, are not available, estimates of number of genes cannot be made, and detection of major genes is not possible. The argument may be presented that genetic correlations are based on too few genotypes. This is only a partial criticism as the number of parents and F_1's can be increased. Other limitations may exist in certain circumstances due to the fact that the parents should be relatively homozygous, considerable range of character expression should be available, and parents should be interfertile so that all F_1's are available.

To counteract some of these shortcomings, particularly the first three, the investigator should include F_2's and backcrosses involving

at least one set of parents. This was done in this study and, for added information, one of the parents contained a marker gene for possible linkage studies.

Considering the problem of estimating gene action in quantitative characters as a whole, probably the field is just beginning to be explored and techniques so far available eventually will be considered crude and ineffective. In the future, statistical and physical attempts, possibly, may be directed at the problem of controlling environmental effects so that single gene study will be feasible. However, such ideas are speculative. Of more immediate concern is the improvement of techniques now available. In this study only two scales of measurement have been used, and these are arithmetic and logarithmic. Possibly a variety of transformations or tests should be available for investigating other general types of interactions.

Breakdown of Yield of Tomatoes into Component Parts

The concept of studying components of complex organizations is not new and has been used in many analytical fields. In the domain of genetics and plant breeding one of the earlier extensive studies was made by Harland (9) in 1919. Yield of cotton was brokendown into morphological entities, and correlations among these components and yield were obtained. Smith (27) in presenting the application of discriminant functions to plant breeding, considered the breakdown of yield in wheat into component parts. Many other studies have been conducted along similar lines with other crops. Maize, in particular, has been subjected to extensive study.

Components of yield of maize, such as ear characteristics have been analyzed and correlated with yield.

Powers (24) probably has conducted the most recent, extensive investigation of components of yield and the relationship of these to yield. This study is of particular importance because it concerns tomatoes, the same experimental material used in this thesis, and involves some of the same components with which this investigation is concerned.

In the present study yield is resolved into sub-characters in the manner illustrated in figure 1.

It may be noted that the breakdown consists essentially of three complete and separate units. Within each unit, the two independent variables bear multiplicative relationships as follows:

Unit 1: yield = (Total number of ripe fruits) x (Average fruit weight)

or: $X_1 = X_2X_3$.

Unit 2: (Total number of ripe fruit) = (Number of clusters) x (Number of fruits per cluster),

or: $X_2 = X_4X_5$.

Unit 3: (Average fruit weight) = (Number of locules) x (Weight per locule),

or: $X_3 = X_6X_7$.

Yield may also be expressed in terms of the four final components directly as follows:

(Yield) = (Number of clusters) x (Number of fruits per cluster) x (Number of locules) x (Weight per locule),

or: $X_1 = X_4X_5X_6X_7$.

To produce an additive nature among these components for the express purpose of obtaining linear relationships, one merely has to take the logarithm of the above equations, i.e., for the last equation,

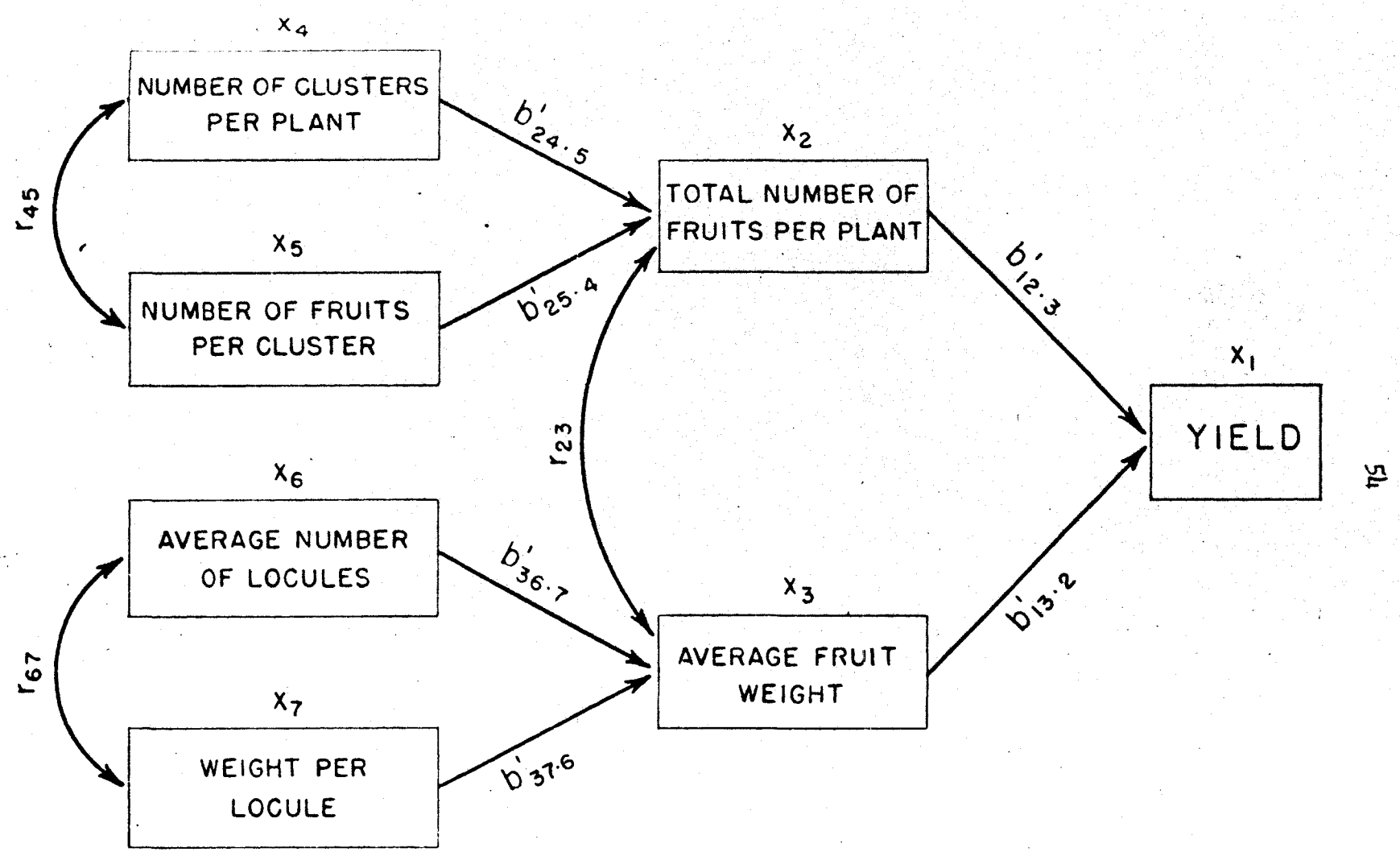


FIGURE 1. DIAGRAMATIC PRESENTATION OF BREAKDOWN OF YIELD INTO COMPONENT CHARACTERS, SHOWING PATH COEFFICIENTS AND CORRELATIONS.

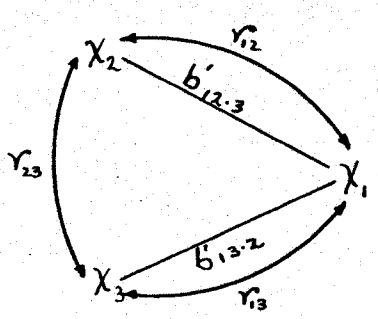
$$\log \hat{X}_1 = \log X_4 + \log X_5 + \log X_6 + \log X_7.$$

Methods of Studying the Relationships among Components

Consider, as an example, the first unit with the variables measured in logarithms so that the following additive relationship holds;

$x_1 = x_2 + x_3$.

The unit represents a "closed circuit" meaning that variation in x_1 is completely determined by variation in x_2 and x_3. Diagramatically the relationships may be presented as follows: (for notation see Snedecor (28).



1. Double arrowed, curved line represents simple correlation i.e. r_{ij}
2. Single arrowed, straight line represents a standard partial regression coefficient i.e. $b'_{i.j}$

With this system, it is a simple matter to determine the portion of variation in x_1, which is attributable to variation in x_2 and x_3, if $r_{23} = 0$. In other words, it is possible to estimate the relative importance of x_2 and x_3 in determining the value of x_1. If $r_{23} \neq 0$, a portion of the variance in x_1 is determined by x_2 and x_3 jointly in a way which cannot logically be ascribed to either or partitioned between them. A cause and effect relationship is assumed as indicated by the straight arrows and with this assumption the standard partial regression coefficients become "path coefficients" as denoted by Sewall Wright (31).

The first step is to estimate the values of the path coefficients from the simple correlations. This is done in the following manner. The model gives rise to two simultaneous equations:

$$r_{12} = b'_{12.3} + r_{23}\, b'_{13.2}$$

$$r_{13} = r_{23}\, b'_{12.3} + b'_{13.2}$$

Solution for path coefficients is as follows:

$$b'_{13.2} = \frac{\begin{vmatrix} 1 & r_{12} \\ r_{23} & r_{13} \end{vmatrix}}{\begin{vmatrix} 1 & r_{13} \\ r_{23} & 1 \end{vmatrix}} = \frac{r_{13} - r_{12} r_{23}}{1 - r_{23}^2} \qquad (1)$$

$$b'_{12.3} = \frac{\begin{vmatrix} 1 & r_{12} \\ r_{23} & r_{13} \end{vmatrix}}{1 - r_{23}^2} = \frac{r_{13} - r_{12} r_{23}}{1 - r_{23}^2} \qquad (2)$$

The next step is to isolate if possible the portion of variation in x_1 which is attributable to each of the independent variables. Two situations should be considered depending on whether $r_{23} = 0$ or $r_{23} \neq 0$. In either case the closed circuit arrangement provides this relationship.

$$(b'_{12.3})^2 + (b'_{13.2})^2 + 2r_{23}\{(b'_{12.3})(b'_{13.2})\} = 1 \qquad (3)$$

Case I. $r_{23} = 0$

When $r_{23} = 0$, then it is obvious that equation (3) reduces to

$$(b'_{12.3})^2 + (b'_{13.2})^2 = 1 \qquad (4)$$

Furthermore, examination of equations (1) and (2) indicate that the standard partial regression coefficients (path coefficients) degenerate to the simple correlations between the two variables. In other words equation

(4) may be expressed simply as

$$(r_{12})^2 + (r_{13})^2 = 1$$

In this way it can be seen that if no correlation exists between the independent variables then the relative importance of each (expressed as a fraction) in determining the dependent variable is merely the square of the simple correlation between it and the dependent variable.

Case II. $r_{23} \neq 0$

When there is a correlation between the two "independent" variables it is not possible to divide the total variance into additive portions each of which measures the effect of one variable, although such an approach is often considered. This point is particularly evident when the correlation (r_{23}) is negative.

The final point of interest concerns the partial correlation coefficients $r_{12.3}$ and $r_{13.2}$. With complete determination of X_1 by X_2 and X_3, both partial correlation coefficients equal the value of one. The sign (plus or minus) is the same as that associated with the corresponding standard partial regression coefficient.

Phenotypic and genotypic correlations

For a two-way classification, as used in this thesis, it is easy to show that the expected mean squares and cross products have the values indicated in Tables 21 and 22. (For analysis of variance alone - see Crump (3)).

The mathematical model is assumed to be the following:

$$y_{klm} = \mu + g_k + r_l + gr_{kl} + \epsilon_{klm} \qquad (7)$$

Where k = 1, 2, . . ., α (number of genotypes)

l = 1, 2, . . ., β (number of replications)

m = 1, 2, . . ., γ (number of plants per plot)

μ = effect common to all observations

g_k = effect common to observations of k^{th} genotype

r_l = effect common to observations of l^{th} replication

gr_{kl} = effect common to observations of both k^{th} and l^{th} classifications

ϵ_{klm} = random effect peculiar to the klm^{th} observation.

Table 21

Analysis of variance of a two-way classification with equal sub-class numbers for i^{th} trait.

Source of variation	dfs	Expectation of mean square
Between varieties	$(\alpha-1)$	$\sigma^2_{\epsilon_{ii}} + \gamma\sigma^2_{\Gamma_{ii}} + \beta\gamma\sigma^2_{G_{ii}}$
Between replications	$(\beta-1)$	$\sigma^2_{\epsilon_{ii}} + \gamma\sigma^2_{\Gamma_{ii}} + \alpha\gamma\sigma^2_{R_{ii}}$
Rep x Var	$(\alpha-1)(\beta-1)$	$\sigma^2_{\epsilon_{ii}} + \gamma\sigma^2_{\Gamma_{ii}}$
Error	$\alpha\beta(\gamma-1)$	$\sigma^2_{\epsilon_{ii}}$

The variance components estimate the variances associated with each of the effects in the math model (7) i.e. $E(g_k^2) = \sigma^2_{G_{ii}}$ for the i^{th} character.

Table 22

Analysis of covariance for a two-way classification with equal sub-class numbers involving traits X_i and X_j .

Source of variation	dfs	Expectation of mean covariances
Between varieties	$(\alpha-1)$	$\sigma^2_{e_{ij}} + \gamma\,\sigma^2_{r_{ij}} + \beta\gamma\,\sigma^2_{c_{ij}}$
Between replications	$(\beta-1)$	$\sigma^2_{e_{ij}} + \gamma\,\sigma^2_{r_{ij}} + \alpha\gamma\,\sigma^2_{R_{ij}}$
Rep x Var	$(\alpha-1)(\beta-1)$	$\sigma^2_{e_{ij}} + \gamma\,\sigma^2_{r_{ij}}$
Error	$\alpha\beta(\gamma-1)$	$\sigma^2_{e_{ij}}$.

The covariance components estimate the covariances between any one effect for one trait and the same effect for another trait, unconfounded by other effects described in the mathematical model. For example, let g_k be, the the k^{th} genotypic effect for yield (X_1), and g'_k; the k^{th} genotypic effect for fruit weight (X_3). Then $E(g_k g'_k) = \sigma^2_{G_{13}}$.

Thus genotypic correlations between any two variables may be estimated merely by use of components of analyses of variances and covariances. The genotypic correlation between the i^{th} and j^{th} variables would be:

$$\hat{r}_{G_{ij}} = \frac{\hat{\sigma}^2_{G_{ij}}}{\sqrt{(\hat{\sigma}^2_{G_{ii}})(\hat{\sigma}^2_{G_{jj}})}}$$

In like manner, correlations may be obtained for any other set of components. Of particular interest are the "environmental" (error) correlations. Consider the mathematical models for two traits measured on the same plant.

Same plant { (Trait x_i - model - $y_{klm} = \mu + g_k + r_l + gr_{kl} + \epsilon_{klm}$
(Trait x_j - model - $y'_{klm} = \mu' + g'_k + r'_l + gr'_{kl} + \epsilon'_{klm}$

ϵ_{klm} is random effect peculiar to the klm^{th} plant and measured on trait x_i. ϵ'_{klm} is similarly described except measured on trait x_j. Then, the correlation:

$$\hat{r}_{E_{ij}} = \frac{\hat{\sigma}^2_{E_{ij}}}{\sqrt{(\hat{\sigma}^2_{E_{ii}})(\hat{\sigma}^2_{E_{jj}})}}$$

measures the association between these two effects. In other words, to what extent does an environmental effect, peculiar to the klm^{th} plant and causing the effect ϵ_{klm} in x_i, tend also to cause the effect ϵ'_{klm} in x_j?

Genotypic correlations may be combined in an appropriate manner to give genotypic path coefficients so that the entire inter-relationships among yield components may be worked out on a genotypic basis alone.

Since these genotypic correlations are based on relatively unrelated parents and non-segregating populations the mathematical model may be written as in equation (7) with no linkage effect and yet with a perfectly valid unconfounded genotypic effect whose variances and covariances may be estimated by components. Thus the genotypic correlations so obtained may be regarded as an estimate of genotypic pleiotropism. This concept may be illustrated by use of mathematical set theory.

Let:

1. The total space correspond to the total array of genes (entire genotype).

2. The set of genes concerned with yield be denoted as $_1G$.

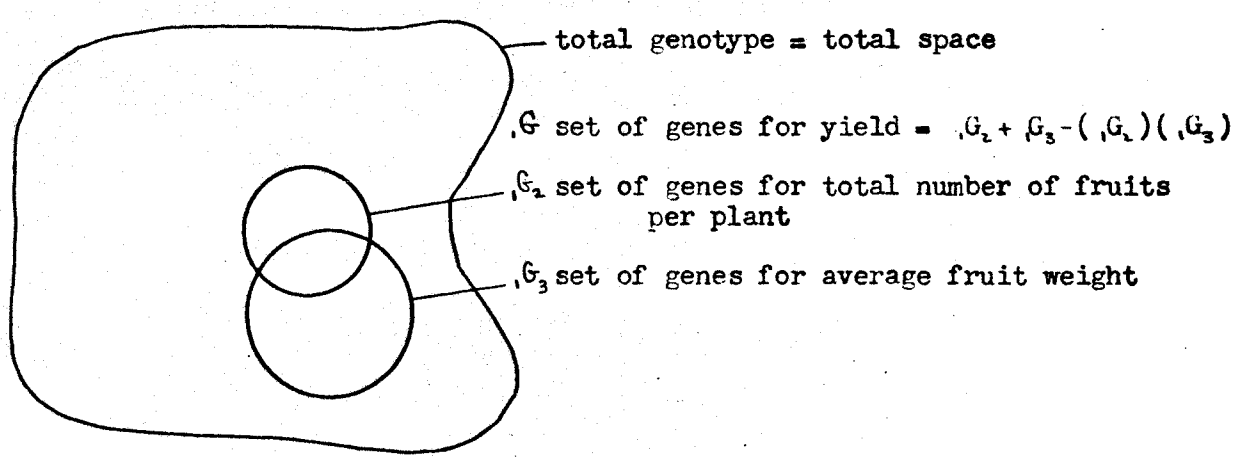
3. The sub-set of genes concerned with variable x_2 be denoted as $_1G_2$.

4. The sub-set of genes concerned with variable x_3 as $_1G_3$.

Symbolize these sets of genes as follows, $(_1G_2)$ and $(_1G_3)$. If there is no genotypic correlation between the two traits, then the two sets would have no genes in common and this would be indicated by disjoint sets $(_1G_2)$ $(_1G_3)$.

If there exists a genotypic correlation, then the two sets would have genes in common and this would be indicated by overlapping sets $(_1G_2)(_1G_3)$.

Diagramatic relationships among the three variables X_1, X_2, and X_3 might appear as follows:

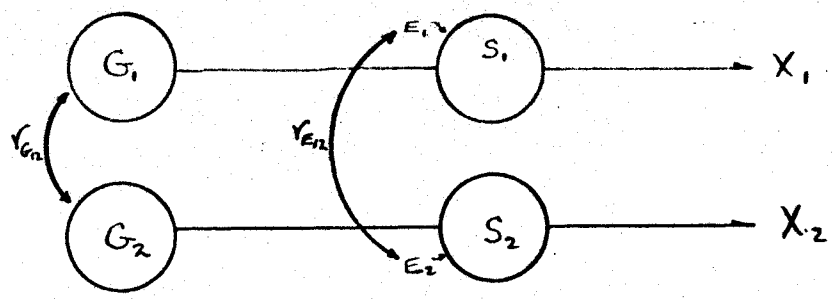


In quite a similar manner the "environmental" (error) correlations might be analyzed to shed some light on the relationships of substrates used by different sets of genes.

Let: (S_1) be the set of substrates on which genes of (G_1) act, and
(S_2) be the set of substrates on which genes of (G_2) act.

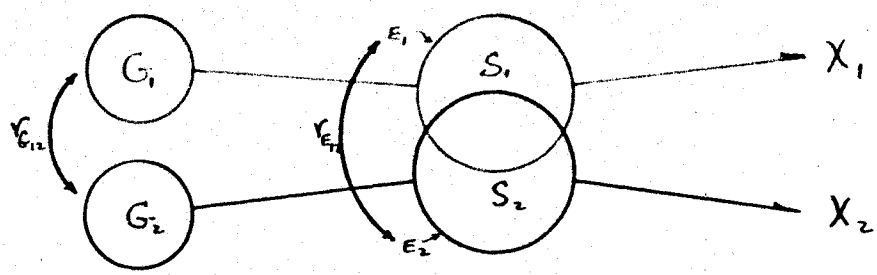
Four general genotype-substrate models may result.

Case I. No genotypic and no environmental correlations

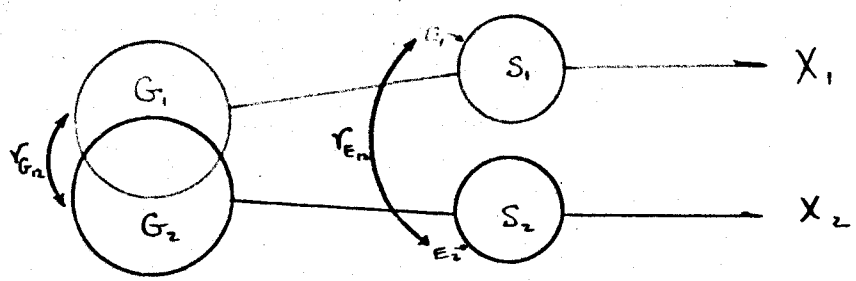


Environmental variations affecting substrate S_1 are independent of environmental variations affecting S_2. One would interpret the above scheme as follows: Gene set G_1 independent of G_2 acts on substrate set S_1 to produce phenotype X_1. Environmental effects E_1 which act on S_1 are independent of environmental effects E_2 which act on S_2.

Case II. No genotypic - with environmental correlations

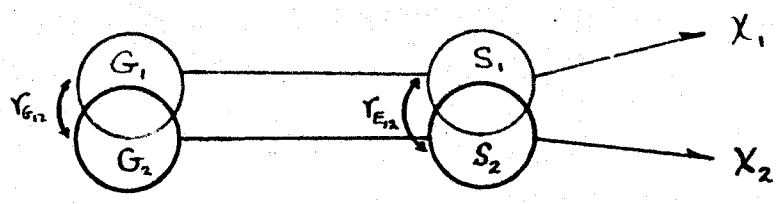


Case III. Genotypic - no environmental correlations



Substrates could be the same but the gene sets act on them at different times.

Case IV. Genotypic and environmental correlations



Further consideration of environmental correlations would suggest that a large negative environmental correlation would indicate competition for a limited, common substrate.

Genotypic correlations may be obtained from segregating populations. Weber (30) used a formula suggested by Hazel (12) and Hazel, et al (13) involving reciprocal regression coefficients between F_2's and F_3's. Such correlations would contain both pleiotropic and linkage effects.

Selection methods for handling components

One of the basic reasons for breaking down a complex character into simpler elements is to direct selection at these component parts, thus causing a greater flexibility in the breeding program. This raises the problem of simultaneous selection for several characters at one time in a manner which will yield maximum amount of genetic gain with a minimum amount of effort.

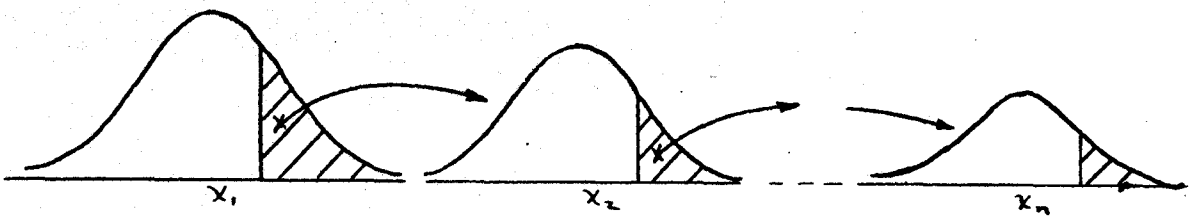
Some of the difficulties and problems which must be considered are; (1) traits are generally complex genetically, (2) sub-traits may be correlated in varying degrees both from a genotypic and phenotypic standpoint, and (3) sub-characters undoubtedly differ in heritability - i.e., some will be highly influenced by environment; others very little. A sound breeding

program should consider all of these problems.

Three general selection schemes may be mentioned.

1. Tandem - Selection is concentrated on one character for enough generations to improve the specified character to the desired level, and then selection is concentrated on another character, etc.

2. Independent culling levels - All characters are subject to selection at the same time. A certain level of desirability is established for each trait. Those plants not measuring up to that level in any one of traits are discarded. In other words, the population passes a gauntlet of tests - one test to a trait; those selected are above the culling level in each of the characters.



3. Selection indices: - All characters are selected simultaneously by an index of net merit;

$$Y = b_1 X_1 + b_2 X_2 + \cdots + b_n X_n$$

where the characters are weighted by b's, so chosen as to give the maximum over-all genetic gain. One basic scheme is that of discriminant functions, originally proposed for breeding techniques by Smith (27).

It is of interest to see how the use of discriminant functions will

take into consideration the relative economic importance of sub-traits, their heritability, and their correlations, (both genotypic and environmental) in combining the components in such a manner as to result in maximum genetic gain for yield. The ensuing discussion follows Smith (27) closely. An attempt is made to simplify the arguments, particularly the statistical approach.

Consider:

1. That there are "n" traits to be combined;

$$\chi_1 , \chi_2 , \cdots , \chi_n$$

2. Each trait, as expressed by the plant, is a result of genotypic and non-genotypic effects;

$$\chi_i = G_i + \epsilon_i$$

3. Components will differ in relative economic importance, and each trait will be weighted accordingly by;

$$a_1 , a_2 , \cdots , a_n$$

4. Actual <u>genotypic</u> value of a plant may be evaluated as follows;

$$H = a_1 G_1 + a_2 G_2 + \cdots + a_n G_n$$

Objective of selection obviously, is to choose plants having the greatest values of H. This is made difficult because one cannot directly evaluate the genotypic values.

5. However, <u>phenotypes</u> of the components can be scored as follows;

$$Y = b_1 \chi_1 + b_2 \chi_2 + \cdots + b_n \chi_n$$

The problem, then, is to discover values for b's such that the function Y may best discriminate those lines which have the greatest genotypic value

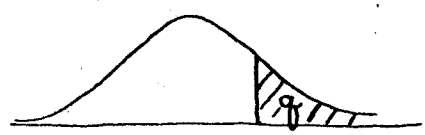
of H. Or; to put it another way, the objective is to find the b's which will give the largest possible correlation between Y and H.

This problem is approached from another point of view. Using the distribution of the phenotypic function $f(Y)$, it is necessary to state in statistical terms the expected genetic gain of the selected group over the original population. Since it is desired to have this as large as possible the expected genetic gain is maximized.

To get the expected genetic gain, consider:

1. Y normally distributed; mean = $\bar{Y}$ variance = V
   then $f(Y) = \frac{1}{\sqrt{2\pi V}} e^{-\frac{1}{2V}(Y-\bar{Y})^2} dY$
2. Transform to unit normal distribution; $n(u; 0, 1)$
   $u = \frac{Y-\bar{Y}}{\sqrt{V}}$; $\therefore Y-\bar{Y} = u(V)^{\frac{1}{2}}$
   $$f(u:0,1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du$$
3. Regression of H on Y is:
   $H-\bar{H} = B(Y-\bar{Y})$ where B is the regression coefficient.
   Since, $Y-\bar{Y} = u(V)^{\frac{1}{2}}$
   then $H-\bar{H} = Bu(V)^{\frac{1}{2}}$.

Select $\frac{1}{q}$ th of Y values;



Expected genetic gain of the selected Y's over the original population will be;

$$E[H-\bar{H}] = \frac{1}{q}\int_{u=u'}^{\infty} B\sqrt{V}\,\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} u\,du$$

$$= \frac{1}{q} B(V)^{\frac{1}{2}} \lim_{t\to\infty} \int_{u=u'}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} u\,du \quad \text{which when integrated}$$

$$= \frac{1}{q} B(V)^{\frac{1}{2}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(u')^2} = \frac{z}{q} B(V)^{\frac{1}{2}}$$

where z is the ordinate of the unit normal curve at u'. $\frac{z}{q} B(V)^{\frac{1}{2}}$ is the expected genetic gain over mean of the original population for a given intensity of selection q. Since the greatest possible genetic gain is wanted, this is the mathematical quantity that must be maximized. Z and q are constants, therefore the maximization process is concerned with the quantity $B(V)^{\frac{1}{2}}$.

The expected genetic gain may be expressed in various ways by different authors:

Smith, Panse: $\frac{z}{q} B(V)^{\frac{1}{2}}$ Hazel $\frac{z}{p} R_{VH} \sigma_H$ Hazel, Lush: $\frac{z}{p} g^2 \sigma_Y^2$

Notation to be used:

$X_i = G_i + \epsilon_i$

V = variance of Y

W = covariance of Y and H

Variances	Covariances
$V_X = t_{ii}$	$X_i X_j = t_{ij}$
$V_G = g_{ii}$	$G_i G_j = g_{ij}$
$V_\epsilon = e_{ii}$	$\epsilon_i \epsilon_j = e_{ij}$

May write:

$$B(V)^{\frac{1}{2}} = \frac{W}{V}\sqrt{V} = \frac{W}{\sqrt{V}} \qquad \text{since } B = \frac{W}{V}$$

$$\log\left(\frac{W}{\sqrt{V}}\right) = [\log W - \tfrac{1}{2}\log V]$$

To maximize with respect to b's:

$$\frac{\partial \log\left(\frac{W}{\sqrt{V}}\right)}{\partial b_i} = 0 \qquad \frac{\partial[\log W - \frac{1}{2}\log V]}{\partial b_i} = \frac{1}{W}\frac{\partial W}{\partial b_i} - \frac{1}{2}\frac{1}{V}\frac{\partial V}{\partial b_i} = 0$$

Proportional equations are:

$$\frac{\partial W}{\partial b_i} = \frac{1}{2}\frac{\partial V}{\partial b_i}$$

For simplicity consider only two characters (n characters considered in summations.)

$H = a_1 G_1 + a_2 G_2$ $\qquad Y = b_1 X_1 + b_2 X_2$

$V = b_1^2 t_{11} + b_2^2 t_{22} + 2 b_1 b_2 t_{12}$ $\qquad = \sum_{i=1}^{n} b_i^2 t_{ii} + \sum_{\substack{i,j \\ i \neq j}}^{n} b_i b_j t_{ij}$

$\frac{1}{2} \frac{\partial V}{\partial b_1} = b_1 t_{11} + b_2 t_{12}$; $\frac{1}{2} \frac{\partial V}{\partial b_2} = b_1 t_{12} + b_2 t_{22}$

For covariance

$W = E[HY]$

$H = a_1 G_1 + a_2 G_2$

$Y = b_1 X_1 + b_2 X_2 = b_1 (G_1 + E_1) + b_2 (G_2 + E_2)$

$HY = a_1 b_1 G_1 (G_1 + E_1) + a_2 b_2 G_2 (G_2 + E_2) + a_1 b_2 G_1 (G_2 + E_2) + a_2 b_1 G_2 (G_1 + E_1)$

then, since $E(G_i E_j) = 0$ for all values of i and j

$W = a_1 b_1 g_{11} + a_2 b_2 g_{22} + a_1 b_2 g_{12} + a_2 b_1 g_{21} = \sum_{i}^{n} a_i b_i g_{ii} + \sum_{\substack{i,j \\ i \neq j}}^{n} a_i b_j g_{ij}$

$\frac{\partial w}{\partial b_1} = a_1 g_{11} + a_2 g_{21}$; $\frac{\partial w}{\partial b_2} = a_2 g_{22} + a_1 g_{12}$ or $\frac{\partial w}{\partial b_i} = \sum_{i} a_i g_{ij}$

From maximizing equation $\left[\frac{\partial w}{\partial b_i} = \frac{1}{2} \frac{\partial v}{\partial b_i} \right]$ we get the following simultaneous equations.

$$b_1 t_{11} + b_2 t_{12} = a_1 g_{11} + a_2 g_{21}$$

$$b_1 t_{12} + b_2 t_{22} = a_1 g_{12} + a_2 g_{22}$$

To solve, set the left-hand side $= \begin{Vmatrix} 1 & 0 \\ 0 & 1 \end{Vmatrix}$, and get the inverse of the variance-covariance matrix i.e., $\| t_{ij} \|^{-1} = \| t^{ij} \|$ or $\| c_{ij} \|$.

Solve for c_{ij}'s

$$c_{11} t_{11} + c_{12} t_{12} = 1 \qquad c_{21} t_{11} + c_{22} t_{12} = 0$$

$$c_{11} t_{12} + c_{12} t_{22} = 0 \qquad c_{21} t_{12} + c_{22} t_{22} = 1$$

Obtain b's

$$b_1 = c_{11} (a_1 g_{11} + a_2 g_{21}) + c_{12} (a_1 g_{12} + a_2 g_{22})$$

$$b_2 = c_{21} (a_1 g_{11} + a_2 g_{21}) + c_{22} (a_1 g_{12} + a_2 g_{22})$$

To find b values - need a_i's, $\| c_{ij} \|$, and $\| g_{ij} \|$.

1. a_i's are assigned to traits according to their economic importance.
2. $\| c_{ij} \|$ is determined merely from the $\| t_{ij} \|$ from following relationship $\| t_{ij} \|^{-1} = \| c_{ij} \|$ where $\| t_{ij} \|$ is the variance-covariance matrix of observable phenotypes.
3. $\| g_{ij} \|$ is obtained simply from components of analysis of variance and covariance.

When b's are found in above manner, the discriminant function $Y = b_1 x_1 + b_2 x_2$ will give a compound score which takes into consideration relative importance of sub-trait, heritability of sub-trait, and genetic correlations. Such a method will yield a compound score which is the linear function most highly correlated with the true genetic value of a variety or line.

From a practical standpoint in plant breeding, obviously, no one method of selection can be universally stamped as best, and probably even in one breeding program a combination of various methods should be used, as indeed usually is done.

Use of discriminant function technique would be time consuming and relatively expensive on an individual plant basis. In animal breeding where relatively small populations are handled and each individual represents a relatively large investment, a careful evaluation of each individual is justifiable. In plants, however, it is relatively easy to increase the number of plants in the initial stages of the selection program and then use independent culling levels to narrow the field down to a selected population. This works best if the early culling on independent levels is done on things, very cheaply observed and acted upon. Harland (11) conducted one of the most extensive selection programs based on the independent culling

levels system in his selection experiments with Peruvian tanguis cotton. His original population consisted of approximately 22,000 single boll samples.

It may be best, in certain circumstances, to grow tremendous numbers of plants in an F_2 and use independent culling levels to decide which individuals shall be included in the selected breeding population. Then in the F_3, which is the first generation to have family structures, a selection index could be used to evaluate different families where closer scrutiny would yield more results.

EXPERIMENTAL MATERIAL

Material used in this study consisted primarily of six inbred lines of tomatoes and all their possible F_1's. These inbreds have been maintained by the Genetics Section, Iowa State College for many years, and appeared exceedingly uniform in this experiment. The following lines were chosen: (1) Red current - a wild species, Lycopersicon pimpinellifolium, small, round, red fruit; (2) Yellow-cherry - L. esculentum, small, round, yellow fruit; (3) Red cherry - L. esculentum, cherry size, red fruit; (4) Goldball - L. esculentum - medium size, round, yellow fruit; (5) Devon - L. esculentum - fairly large, round, red fruit, (this parent has the "u" marker gene, the presence of which prevents the top part of the fruit from becoming dark green; instead the fruit has a uniformly light green color during growth before ripening); (6) Matchless - L. esculentum, large, somewhat oblate, red fruit. F_2 and both backcrosses for the cross Devon x Matchless were also included in the study.

The parental numbers as attached to the above lines will henceforth be used to denote the specific inbred lines.

Crossing was done during the winters of 1946 and 1947 under greenhouse conditions. On May 2, 1947, the final experiment was started by planting all seeds in greenhouse flats. Fifteen days later the seedlings were potted in two-inch pots. All greenhouse material was randomized in the same order as in the field planting arrangement. On the 15th and 16th of June, transplantation of the plants to the field was accomplished. The growing season was wet at the beginning but rather dry during the last few months of the summer - apparently excellent conditions for tomato growth. Out of approximately 1,200 plants, only a few replants were necessary, and at the end of the experiment only three plants had died.

The experimental design consisted of six randomized blocks. Within each block were the following: (1) one plot of each inbred and F_1, (2) two plots of each backcross, and (3) three plots of F_2. Each plot consisted of seven plants and all plots were randomized with the restriction that plots of any one segregating population would be adjacent. Plots were approximately three and one-half feet apart and plants were about the same distance apart within a plot. The entire experiment was surrounded by border plants.

All sampling techniques were based on individual plants. For some characteristics a complete sampling scheme was possible; for others a systematic sub-sampling method was used. With this system a random choice was made of the second or third plant in the plot, and then beginning with this plant, every other plant was chosen. Thus three plants out of the seven were sampled from each plot. All segregating plots were completely sampled. A complete sample was taken for the following traits; (1)

flowering date, (2) number of days from flowering to fruit ripe, (3) number of flowers per cluster, (4) number of fruits per cluster, and (5) number of locules per fruit. Those for which sub-sampling methods were used include; (1) yield, (2) total number of fruits per plant, (3) number of clusters having ripe fruit per plant, (4) average fruit weight and (5) average weight per locule.

Flowering date and maturity time were taken from daily observations. Data on number of flowers and fruits per cluster were obtained from the first three clusters which were marked by variously colored tags. The character, number of locules per fruit, was based on a random ten fruit sample taken from each plant fairly late in the harvesting season. Yield represented total ripe fruit. Ripe fruits were harvested from plants throughout the season at intervals of approximately every two weeks. This procedure was used to minimize loss due to rotting, of which there was very little. Fruits were not picked until they had turned to a reddish flush (or yellowish color). Final harvest was made on September 28, 1947 for the first three replications, and during the next three days the last three replications were harvested. Two light frosts had previously killed the foliage but had not damaged the fruits.

Total number of ripe fruit per plant was obtained by actual count at the time fruits were weighed. For the small sizes counting trays were used. Estimate of number of clusters having ripe fruit was obtained by dividing total number of fruits by number of fruits per cluster. Average fruit weight was determined by dividing total plant yield by total number of fruit. Finally, average weight per locule was obtained by dividing average fruit weight by number of locules per fruit.

Ranges in character expression were exceedingly large among the parents, as can be seen from Table 23. The following abbreviated Table illustrates this point.

Table 23

Ranges of character expression found among the parental lines (arithmetic values)[1]

	Low parent	High parent
x_1 (Yield)	678	2253
x_2 (Total number of fruit)	16	1287
x_3 (Av. fruit weight)	.5	142.6
x_4 (Number of clusters)	6.1	133.5
x_5 (Number of fruits per cluster)	2.6	10.5
x_6 (Number of locules	2.0	6.8
x_7 (Weight per locule)	.3	20.5
x_8 (Flowering date)[2]	5.0	20.4
x_9 (Maturity time)[3]	38.3	44.9

1 Weights are in grams

2 Flowering date values are coded with first day corresponding to the first flower occurring in the experiment.

3 Maturity time represents the number of days from flower to ripe fruit.

To give an idea of the amount of material analyzed, it was estimated that between 75,000 and 85,000 fruits were harvested, weighed and counted.

ANALYSIS OF EXPERIMENTAL RESULTS

Estimation of Average Gene Action

In presenting the data, the problem of estimating gene action is considered first because it is with this phase that the thesis is primarily concerned. The problems of inter-relationships between components and selection indices, will be taken up later.

Before analyzing the experimental data, it is necessary to consider a few points that pertain to the analysis of the majority of the characters.

All analyses of variances and covariances of yield components were based on log data. In other words, the individual plant values for traits X_1, X_2, X_3, X_4, X_5, X_6, and X_7 were all transformed to three place logarithms for calculating procedures. This was done for at least two reasons. First, linear relationships among the components were obtained with log transformation. Second, with the extreme ranges for all the characteristics, the means and variances were correlated. Log transformation would tend to rectify this situation.

In the process of choosing between logarithmic and arithmetic gene action, it was necessary to compare analyses based on arithmetic means with those based on logarithmic means. This created a problem of estimating the standard errors for the means on both scales of measurement. Obviously this involves an approximation process. The regular analysis of

variance will yield an estimate of one, which is, in most cases the logarithmic error. Then, the arithmetic error must be approximated from the log error. The problem can be approached in two ways. One way is by use of a transformation to obtain an approximate estimate.

Case I - Obtaining approximate variances for arithmetic values when variances of the log values are known.

Given: $y = \log_{10} x$; $x = 10^y$ (log to base 10 used)

Know: V_y

Want: V_x

$$\frac{dx}{dy} = 10^y \log_e 10 \; ; \quad dx = (10^y)(\log_e 10)\, dy$$

$$E(dx)^2 = E\{10^{2y}(\log_e 10)^2 (dy)^2\}$$

$$V_x \approx 10^{2y}(\log_e 10)^2 V_y$$

$$V_x \approx (5.30192)(10^{2y}) V_y$$

For different constant parent groups, the corresponding $\bar{y}$ for each group was used.

Case II - Obtaining approximate variances of log values when variances of the arithmetic values are known.

Given: $y = \log_{10} x$; $y = \left(\frac{1}{\log_e 10}\right) \log_e x$

Know: V_x

Want: V_y

$$\frac{dy}{dx} = \left(\frac{1}{\log_e 10}\right)\frac{1}{x} \; ; \quad dy = \left(\frac{1}{\log_e 10}\right)\frac{dx}{x}$$

$$E(dy)^2 = \left(\frac{1}{\log_e 10}\right)^2 \frac{E(dx)^2}{E(x)^2}$$

$$V_y \approx \left(\frac{1}{\log_e 10}\right)^2 \frac{V_x}{\bar{x}^2}$$

The other method of obtaining both arithmetic and logarithmic standard errors is to run analyses of variances in both scales, but, obviously, this procedure warrants theoretical criticism. To partially rectify this situation, in which, for example, the data are distributed logarithmically, a separate analysis of variance for arithmetic data would be calculated for each constant parent group. (See analysis for fruit size for this method.)

Another point which pertains to almost all analyses is the question of whether to use the logs of arithmetic means or the logarithmic means themselves, which are consistently the smaller value of the two. Which method is used does not appear to change the analyses to any extent. Both schemes were tried on fruit weight and were found to yield essentially, the same conclusions. So nearly alike were the analyses that both yielded the same percentages for the regression components. Since the data are usually recorded in an arithmetic scale in plant breeding problems, the comparison in general must be made using arithmetic means with log transformations of these means. Therefore it was decided to do the same in this thesis.

One final general technique should be mentioned. For each of the traits, one cross (5x6) is carried to the segregating populations. This entails a study of the frequency distributions for each of the generations, P_1's, F_1, F_2, B_1, and B_2. In some of the characters, considerable replication effect is discernible, giving rise to variation which may be eliminated.

Consider one non-segregating line at a time in six replications.

Math model $X_{ij} = \mu + r_i + \epsilon_{ij}$ where r_i = replication effect

ϵ_{ij} = individual error

By calculating $\bar{X}_{i.}$ (mean of i^{th} rep) for each replication and determining $\bar{X}_{i.} - \bar{X}_{..}$ ($\bar{X}_{..}$ is over-all mean), the individual items may be adjusted so as to remove r_i , since $r_i = \bar{X}_{i.} - \bar{X}_{..}$.

However, with segregating populations, this procedure may cause an unknown amount of confounding of genetic and replication effects.

Use of the least squares solutions for tests of b_i was not attempted for the various traits studied. The c.p.r. trends are only a part of the evidence leading to an estimation of gene action and thus such an omission is not too serious.

X_1 - Yield

Arithmetic means for both parents and F_1's are given in Table X_1-1 along with the c.p.r. trend. It may be noted that only one F_1 did not outyield the highest parent. This F_1 resulted from the cross (2x5). Amount of heterosis can be noted in Table X_1-3 where average "h" values are listed for each C.P. (constant parent) group. This phenotypic "h" value measures the increase of F_1 over the midparent values relative to the difference between highest parent and midparental value. Obviously, h = 1 is complete dominance, and the term heterosis will be applied to those cases when h > 1 or h<-1 or when the F_1 value exceeds either parental value. (The terms heterosis and hybred vigor will be considered as synonymous throughout this thesis). The over-all value of h = +2.41 indicates considerable heterosis for the F_1's in general.

The c.p.r. trend is consistently downward going from practically +1.00 to 0 with increase of parental values. This would indicate that the best

model would be arithmetically cumulative gene action with at least complete dominance. Geometric gene action is ruled out because the regression trend is directly opposite to such a hypothesis. However the log Table X_1-2 is included to give a table having means with a common least significant difference.

Going to the Table X_2-3 one can see that four of the six "deviations from regression mean squares" are significantly greater than the error mean square. Three of these are highly significant. Interpretation of the components E, D and B, is a bit confused, in that, with C.P.s 4, 5 and 6 there is actually very little variation among the F_1's within each group. However, whether one considers only the first three C.P. groups, or all six, it is evident that there is considerable variation attributable to deviations from regression.

The evidence points to a model with arithmetic gene action and dominance on the order of complete dominance to over- overdominance. The "h" values indicate over-dominance, but the regression trend only goes as far as complete dominance.

It is of particular interest to see how the "h" values of yield are determined by dominance values of component parts. Richey (26) has pointed out how heterosis may result from "mock dominance" which in reality is an interaction between sub-traits. Powers (24) also pointed out how heterosis in yield can result from the multiplicative nature of immediate components when components exhibit only partial dominance. It is possible to examine this condition more critically and see how the dominance values of number of fruit (X_2) and average fruit weight (X_3) determine the dominance value of yield (X_1). It is necessary to look

Table X_1-1

Arithmetic mean P_1 and F_1 values for yield of ripe fruit (X_1). Average F_1 value for each constant parent group is given as well as c.p.r. coefficients.

Parents → / Parents ↓	1	2	3	4	5	6
1	678	1198	1681	2523	2552	2456
2	1198	921	1437	1824	1953	2320
3	1681	1437	1144	2539	2409	2265
4	2523	1824	2539	1628	2683	2341
5	2552	1953	2409	2683	1996	2452
6	2456	2320	2265	2341	2452	2253
C.P. $\bar{F}_1$ =	2082	1746	2066	2382	2410	2367
c.p.r. coefficients	.979	.681	.577	.142	.137	.025

Table X_1-2

Logarithmic mean P_1 and F_1 values for yield of ripe fruit (X_1)

	1	2	3	4	5	6
1	2.8312	3.0785	3.2256	3.4019	3.4069	3.3902
2		2.9643	3.1575	3.2610	3.2907	3.3655
3			3.0584	3.4047	3.3818	3.3551
4				3.2117	3.4286	3.3694
5					3.3002	3.3895
6		L.S.D. $= t_{.05}\hat{\sigma}_{\bar{d}} = .0931$				3.3528

Table X_1-3

Summary of statistics used in estimating gene action. Analyses on arithmetic data.

C.P.	C.P. value	c.p.r.	Components of regression A of V % E	D	B	Regression M.Sq.	$\Sigma d_{\bar{y}\cdot x}^2$ M.Sq.		Ave "h" value for each c.p.
1	687	+.98	1.8	6.0	92.2	1,199,808**	99,235**	22,226	+2.51
2	921	+.68	2.1	0	97.9	752,257**	7,312	16,103	+2.10
3	1,144	+.58	3.2	11.8	85.0	611,945**	104,300**	22,673	+2.94
4	1,628	+.14	21.6	73.4	5.0	38,305	136,624**	31,036	+3.04
5	1,996	+.14	34.8	65.2	0	29,387	91,973*	31,961	+2.40
6	2,253	+.03	100.0	0	0	734	9,234	31,153	+1.44
			27.3	26.1	46.6				h =+2.41
	b_2 = -.0006								

Notation used:

1. C.P. = constant parent.

   c.p.r. = constant parent regression.

   E = error component in regression analysis of variance.

   D = deviations from regression component in regression A of V.

   B = component estimating variance attributable to regression in regression A of V.

   $\hat{\sigma}^2_{\bar{F}}$ = variance associated with the F_1 means tested for each c.p. group separately.

   $h = \frac{F_1 - M_P}{P_2 - M_P}$.

ahead to the gene estimation of X_2 and X_3 to see that both types of gene action are assumed to be on logarithmic basis.

Since $X_1 = X_2X_3$

$\log X_1 = \log X_2 + \log X_3$

Consider all values for the rest of the discussion to be entirely log values.

Then; $F_1 = M_P + h\,(P_2 - M_P)$ or $h = \frac{F_1 - M_P}{P_2 - M_P}$

Notation to be used:

1. For yield values, "y" will be placed in front of the various statistics i.e., $_yF_1$, $_yM_P$, $_yh$, and $_yP_2$.
2. For number of fruits, "n" will be placed in front of statistics, i.e., $_nF_1$, $_nM_P$, etc.
3. For fruit weight, "w" will be used, i.e., $_wF_1$, $_wM_P$, etc.

P_2 will designate the highest yielding parent, P_1 the lowest. For yield in tomatoes in this study, $_yP_2$ always has fewer but heavier fruits than $_yP_1$, i.e., $_nP_2 < {}_nP_1$; $_wP_2 > {}_wP_1$.

Now: $_yF_1 = {}_nF_1 + {}_wF_1$

Since, $_nF_1 = {}_nM_P + {}_nh\,(_nP_1 - {}_nM_P)$

and, $_wF_1 = {}_wM_P + {}_wh\,(_wP_2 - {}_wM_P)$

then, $_yF_1 = {}_nM_P + {}_nh\,(_nP_1 - {}_nM_P) + {}_wM_P + {}_wh\,(_wP_2 - {}_wM_P)$

$_yh = \frac{_yF_1 - {}_yM_P}{_yP_2 - {}_yM_P}$; $_nM_P = \frac{_nP_1 + {}_nP_2}{2}$; $_wM_P = \frac{_wP_2 + {}_wP_1}{2}$

$$_yh = \frac{\frac{_nP_1 + {}_nP_2}{2} + {}_nh\left(_nP_1 - \frac{_nP_1 + {}_nP_2}{2}\right) + \frac{_wP_1 + {}_wP_2}{2} + {}_wh\left(_wP_2 - \frac{_wP_1 + {}_wP_2}{2}\right) - \frac{_yP_1 + {}_yP_2}{2}}{_yP_2 - \frac{_yP_1 + {}_yP_2}{2}}$$

then $$_yh = \frac{{}_nh\,(_nP_1 - {}_nP_2) + {}_wh\,(_wP_2 - {}_wP_1)}{{}_nP_2 - {}_nP_1 + {}_wP_2 - {}_wP_1} \qquad (1)$$

The next step is to determine the relationship between number of fruit and size of fruit. On examining figure 2 it can be noted that an almost exact linear relationship exists. A change in number of fruits from one parent to another is accompanied by a negative change in average fruit weight. The regression value of .78 is approximately equal to the ratio $\frac{7}{9} = .777\ldots$, therefore the increment values of seven and nine may be substituted in formula (1) to give

$$yh = \frac{nh(7) + wh(9)}{2} .$$

Table X_1-4 may be devised to show how dominance values for traits "number of fruits" and "weight per fruit" combine to give dominance values for yield. Examination of the Table X_1-4 points out that there are relatively few combinations which will <u>not</u> give heterosis (yh values greater than 1). Positive heterosis can come about even when one trait exhibits negative dominance values if the other has positive "h" values near .8 or higher.

The question of a tester for general combining ability may be examined. C.P. 1 (red currant) with its F_1's was the only inbred that gave a direct and accurate measure of the general combining ability. The greatest specific combining ability is obtained in all cases but one, when consecutive parents are crossed, i.e., 1x2, 2x3, etc.

In considering the segregating populations C.P.'s 5 and 6 were quite similar in yield so that their frequency distributions overlapped. (See Table 37). Means and"h" values are listed in Table X_1-5 for the arithmetic values.

The "h" values for F_2, B_1, and B_2 do not closely agree, and B_2 is the only segregating population exhibiting a "h" value greater than one. In

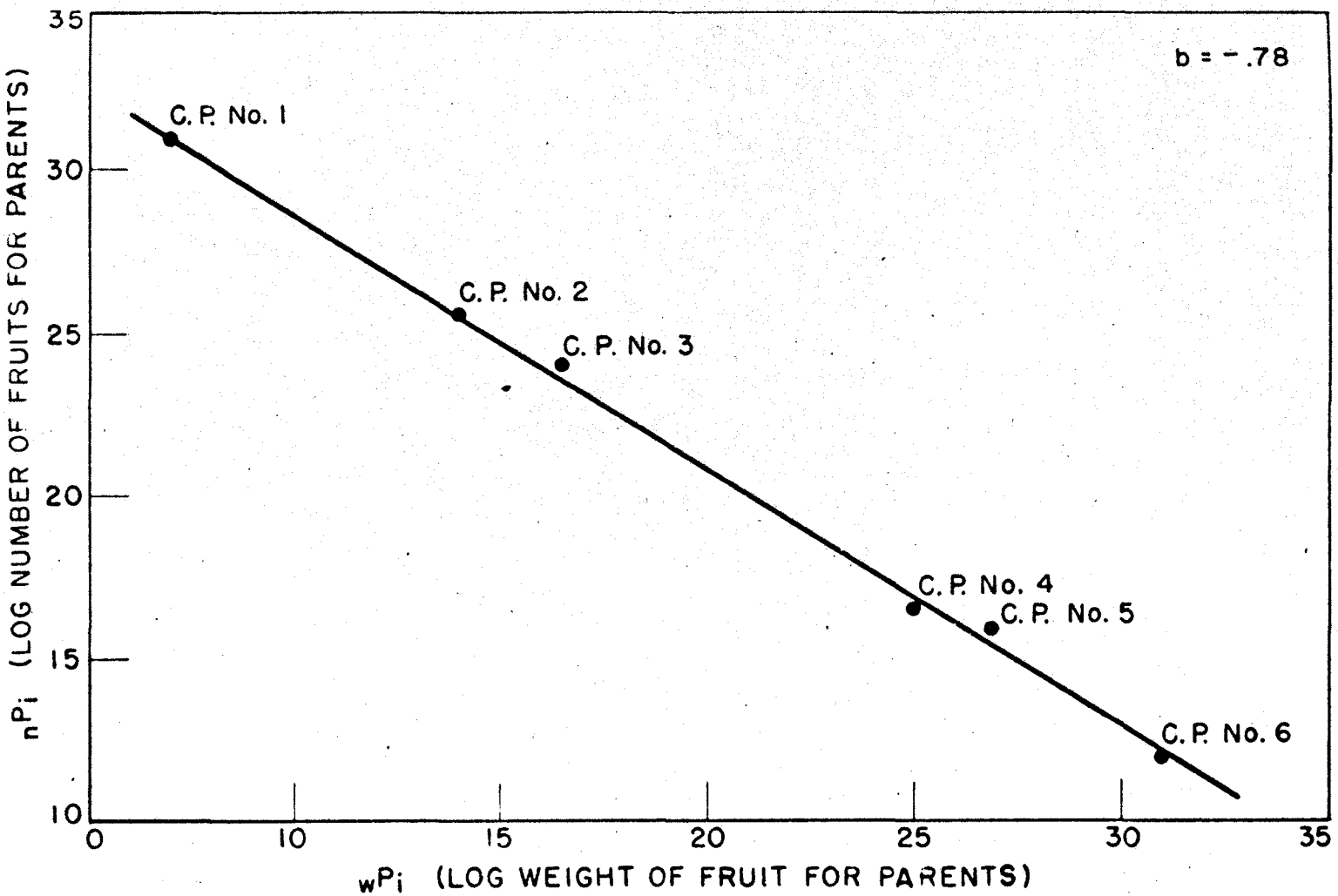


FIGURE 2. LOGARITHMIC VALUES OF TOTAL NUMBER OF FRUIT PER PLANT ($_nP_i$) AND AVERAGE FRUIT WEIGHT ($_wP_i$) PLOTTED FOR EACH CONSTANT PARENT

Table X,-4

Theoretical "h" values for F_1 yields when considering different h values for component parts - number of fruits and weight per fruit.

values for wh ↓	values for nh										
	-1.0	-.8	-.6	-.4	-.2	0	+.2	+.4	+.6	+.8	+1.0
-1.0	-8.0	-7.3	-6.6	-5.9	-5.2	-4.5	-3.8	-3.1	-2.4	-1.7	-1.0
-.8	-7.1	-6.4	-5.7	-5.0	-4.3	-3.6	-2.9	-2.2	-1.5	-.8	-.1
-.6	-6.2	-5.5	-4.8	-4.1	-3.4	-2.7	-2.0	-1.3	-.6	+.1	+.8
-.4	-5.3	-4.6	-3.9	-3.2	-2.5	-1.8	-1.1	-.4	+.3	+1.0	+1.7
-.2	-4.4	-3.7	-3.0	-2.3	-1.6	-.9	-.2	+.5	+1.2	+1.9	+2.6
0	-3.5	-2.8	-2.1	-1.4	-.7	0	+.7	+1.4	+2.1	+2.8	+3.5
+.2	-2.6	-1.9	-1.2	-.5	+.2	+.9	+1.6	+2.3	+3.0	+3.7	+4.4
+.4	-1.7	-1.0	-.3	+.4	+1.1	+1.8	+2.5	+3.2	+3.9	+4.6	+5.3
+.6	-.8	-.1	+.6	+1.3	+2.0	+2.7	+3.4	+4.1	+4.8	+5.5	+6.2
+.8	+.1	+.8	+1.5	+2.2	+2.9	+3.6	+4.3	+5.0	+5.7	+6.4	+7.1
+1.0	+1.0	+1.7	+2.4	+3.1	+3.8	+4.5	+5.2	+5.9	+6.6	+7.3	+8.0

Formula used was $yh = \frac{nh(7) + wh(9)}{2}$

constructing a gene model the "h" value used is the average "h" for the entire $P_1 - F_1$ table instead of the average from the segregating populations. The reason is that the $P_1 - F_1$ data consider six different parents which collectively give a large range of character expression, whereas the segregating populations trace back to only the top two parents which differ little in yield. Therefore statistics from the $P_1 - F_1$ data should carry more weight than those from the segregating populations.

Table X_1-5

Means and "h" values for various generations concerned with the segregating populations. Arithmetic data only.

	P_1	P_2	F_1	F_2	B_1	B_2	
mean	1996	2253	2452	2135	2121	2375	A. "h"
"h"			2.55	.16	.95	2.90	1.64

X_2 - Total number of ripe fruits per plant

An examination of Table X_2-1 or the "h" Table X_2-4 leads to the fact that the F_1 values are generally intermediate between the parents. Only two F_1's exceeded the highest parent.

The c.p.r. analysis on arithmetic values yields an increasing trend with increasing parental values. This indicates either arithmetic gene action with negative dominance or an interaction, probably of a logarithmic nature. Two immediate pieces of evidence against arithmetic action with negative dominance are, (1) the highest c.p.r. has value of +2.02, which is beyond the limit of arithmetic scheme (barring negative over-dominance

which is not present in this case); (2) heterosis, when found, is in a positive direction, not negative.

The next procedure, in a more detailed analysis, is to transform the arithmetic means to logs and compare the arithmetic with the logarithmic analysis. With log data, Table X_2-2, the c.p.r. trend is completely reversed indicating that considerable positive dominance must be associated with a log scale of measurement. It should be mentioned in passing that both c.p.r. trends are consistently uniform, and are severe trends indicating that both have genetic interpretation of considerable importance.

On examining the "h" Table X_2-4, one may note that the arithmetic values are highly erratic in general. By log transformation the "h" values become much more uniform. In arithmetic values it is interesting to notice how the dominance appears to change uniformly from high positive to negative values as C. P. 1 is crossed with parents of decreasing values. This apparent change in dominance can be explained, largely, by assuming logarithmic gene action because these same F_1's give relatively uniform results with transformed data.

The regression analysis of variance, as indicated in Table X_2-3, does not contribute much to gene action estimation except that in log data the four out of six highly significant values for "deviations from regression mean square" are consistent with the hypothesis of log action and considerable positive dominance. It is believed that one of the reasons that no more than three "deviations from regression mean square" are significant in arithmetic analysis is due to large values of $\hat{\sigma}^2_{\bar{F}_1}$. It appears that the approximation method of estimating $\hat{\sigma}^2_{\bar{F}_1}$ yields values

which are consistently large.

The segregating populations, give results, as illustrated in Table X_2-5, which agree closely with the rest of this analysis.

Table X_2-5

Means and "h" values for various generations concerned with the segregating populations.

		P_1	P_2	F_1	F_2	B_1	B_2	
Means	Arith	15.9	40.6	36.3	32.9	23.2	40.4	
	Log	1.20140	1.60853	1.55630	1.51720	1.36549	1.60638	Av."h"
"h" values	Arith			.60	.75	.18	.97	.63
	Log			.72	1.10	.61	.98	.85

These average "h" values compare very well with the over-all F_1 averages for arithmetic and logarithmic data.

In summarizing, the various sources of evidence point consistently to a gene model involving logarithmic gene action with considerable positive dominance, i.e., "h" $\gtrsim +.80$. The evidence for log action consists first and most important, in the increasing c.p.r. trend. Second, the final c.p.r. value equals + 2.02 which exceeds +1.0. Third, the table of "h" values tends to become more uniform on log transformation.

X_3 - Average fruit weight

From either the P_1-F_1 Tables X_3-1 and X_3-2 or the "h" Table X_3-4, it is obvious that the F_1 values are all intermediate between the two parental

Table X_2-1

Arithmetic mean P_1 and F_1 values for "total number of ripe fruit (X_2) constant parent regressions coefficients are also listed.

Constant Parents	1	2	3	4	5	6	
1	1287	1187	1201	661	612	561	
2		358	442	226	188	155	
3			260	212	193	139	
4				50	60	32	
5					41	36	
6						16	
c.p.r. coefficients	2.02	.79	.82	.48	.45	.42	$b_2 = +.0012$

Table X_2-2

Logarithmic mean P_1 and F_1 values for total number of fruits per plant (X_2). Also constant parent regression coefficient.

Constant Parents	1	2	3	4	5	6	
1	3.1096	3.0745	3.0795	2.8202	2.7868	2.7490	
2		2.5539	2.6454	2.3541	2.2742	2.1903	
3			2.4150	2.3263	2.2856	2.1430	
4				1.6990	1.7782	1.5052	
5					1.6128	1.5563	
6	L.S.D. $= t_{05}\,\hat{\sigma}_{\bar{d}} = .0877$					1.2041	
c.p.r. coefficients	.28	.47	.48	.68	.64	.82	$b_2 = -.2642$

Table X_1-3

Summary of statistics used in estimating gene action for trait X_2. Both arithmetic and logarithmic analysis presented.

C.P.	C.P. value	c.p.r.	Components of regression A of V in % E	D	B	Regression M.S.	$\sum d_{1x}^2$ M.S.	$\hat{\sigma}_{\bar{E}}^2$	Ave. "h" for each C.P.	Ave. "h" disregarding 2 erratic values
a. Arithmetic data										
1	1287	2.02	.8	1.3	97.9	388,274**	8,075	3,153	.28	.28
2	358	.79	.1	.2	99.7	743,213**	1,775*	513	.67	.17
3	260	.82	.1	0	99.9	781,576**	370	486	.90	.44
4	50	.48	0	.3	99.7	251,673**	732**	102	.77	.15
5	41	.44	0	.3	99.7	213,229**	540**	93	.81	.21
6	16	.42	0	.1	99.9	189,643**	119	57	.04	.04
	b_1 = +.0012		.2	.4	99.4				+.58	+.22
b. Logarithmic data										
1	3.1096	.28	.9	.7	98.4	.099814**	.001607	.000934	.71	.71
2	2.5539	.47	.2	.5	99.3	.507812**	.003757**	.000934	.92	.57
3	2.4150	.48	.2	.8	99.0	.543771**	.005565**	.000934	1.04	.72
4	1.6990	.68	.1	.1	99.8	1.071184**	.001498	.000934	.99	.52
5	1.6128	.64	.1	.5	99.4	.911775**	.005773**	.000934	1.04	.60
6	1.2041	.82	.1	.3	99.6	1.041881**	.004448**	.000934	.51	.51
	b_1 = -.2642		.3	.5	99.2				+.87	+.61

Table X_2-4

"h" values for F_1's. Arithmetic and logarithmic data.

a. Arithmetic

	1	2	3	4	5	6	
1		+.78	+.83	-.01	-.08	-.14	
2			+2.71	+.14	-.07	-.19	
3				+.54	+.39	+.01	
4					+3.22	-.06	
5						+.60	
6							
Ave. "h" values	+.28	+.67	+.90	+.77	+.81	+.04	$\bar{h}$ +.58
disregarding (h=2.71) (h=3.22)	+.28	+.17	+.44	+.15	+.21	+.04	+.22

b. Logarithmic

	1	2	3	4	5	6	
1		+.87	+.91	+.59	+.57	+.62	
2			+2.32	+.53	+.41	+.46	
3				+.75	+.68	+.55	
4					+2.84	+.22	
5						+.72	
6							
Ave h for c.p.	+.71	+.92	+1.04	+.99	+1.04	+.51	$\bar{h}$ +.87
disregarding (h=2.32) (h=2.84)	+.71	+.57	+.72	+.52	+.60	+.51	+.61

means. In fact the F_1's are much nearer the low parent in all cases. By glancing over the arithmetic Table X_3-1 one can see why there has been considerable controversy over the choice of either, (1) arithmetic gene action with considerable negative dominance, or (2) logarithmic gene action with little or no dominance. Considerable evidence will be presented here in an attempt to differentiate between the two so that the most appropriate model may be chosen.

First consider the c.p.r. analysis on arithmetic data. The c.p.r. trend is uniformly increasing with the highest value exceeding one. Such an increasing trend indicates either one of the aforementioned models and as such does not help differentiate the two. However, the first bit of evidence is that the final c.p.r. value exceeds one, and by following the argument given for X_2 this suggests logarithmic gene action.

The next step is to observe the change in analysis due to logarithmic transformation of arithmetic means. The severe c.p.r. trend found in arithmetic values is drastically diminished to a slight, fairly uniform, increasing trend. This in itself would indicate that on logarithmic scale there occurs a consistent low negative dominance effect.

Examination of regression analysis of variance yields important information on the problem. In the arithmetic analysis all "deviations from regression mean square" are highly significant. The significance of these mean squares drops considerably in the transformed data. This indicates that considerable metrical bias effect has entered into the arithmetic mean square, since it is removed on transformation.

In the component analysis the same situation is found. In arithmetic values, component D averages 8% of the variance, whereas most of this

disappears in transformation. This indicates that most of non-additive gene effects (estimated by D), in arithmetic values, is due to logarithmic gene action.

Considering the F_1 "h" Tables X_3-4, the fairly erratic arithmetic "h's" are smoothed out considerably in the log "h's" except for two erratic values. "h" statistics from the segregating populations are given in Table X_3-5 and again erratic arithmetic estimates are made more uniform with log transformations.

Table X_3-5

Means and "h" values for various generations concerned with segregating populations. (Log values coded.)

		P_1	P_2	F_1	F_2	B_1	B_2	
Means:	Arith	49.1	142.6	67.3	67,8	58.7	93.5	
	Log	1.6911	2.1541	1.8280	1.8312	1.7686	1.9708	Av. "h"
"h"	Arith			-.61	-1.20	-.59	-1.10	-.88
	Log			-.41	-.79	-.33	-.58	-.53

These average "h" values are consistently higher than the F_1 "h's", and it should be remembered that the F_1 values extend over a tremendous range of fruit sizes involving six different parents, whereas the segregating population analysis is based on only the two top parents. Therefore more weight should be given to the $P_1 - F_1$ technique in determining the relative amount of dominance. The evidence of low amount of dominance as indicated by the average $P_1 - F_1$ value $\bar{h} = -.07$ is further verified by the slight negative dominance trend in c.p.r. coefficients. This joint evidence

Table X_3-1

Arithmetic mean P_1 and F_1 values for "Average fruit weight" (X_3) constant parent regression coefficients are also listed.

Constant Parents	1	2	3	4	5	6
1	.5	1.0	1.4	3.8	4.2	4.4
2		2.6	3.2	8.1	10.3	14.9
3			4.4	12.0	12.5	16.3
4				32.7	45.0	73.2
5					49.1	67.3
6						142.6
c.p.r. coefficients	.02	.09	.10	.48	.41	1.39

Table X_3-2

Logarithmic mean P_1 and F_1 values for "Average fruit weight" (X_3) constant parent regression coefficients are also listed. Log values coded (1 added to mantissa in order to avoid negative logarithms)

Constant Parents	1	2	3	4	5	6
1	.7160	1.0086	1.1492	1.5832	1.6233	1.6425
2		1.4082	1.5092	1.9063	2.0141	2.1741
3			1.6405	2.0799	2.0962	2.2117
4				2.5146	2.6530	2.8643
5					2.6911	2.8278
6						3.1541
c.p.r. coefficients	.39	.48	.45	.54	.51	.62

L.S.D. $= t_{.05}\,\hat{\sigma}_d = .1080$

Table X_3-3

Summary of statistics used in estimating gene action for trait X_3. Both arithmetic and logarithmic data presented.

C.P.	C.P. value	c.p.r.	Components of regression A of V in % E	D	B	Regression M.S.	$\Sigma d_{\bar{y}\cdot x}^2$ M.S.	$\hat{\sigma}_{\bar{E}}^2$ *	Ave. "h" for each C.P.
a. Arithmetic data									
1	.5	.02	.1	19.9	80.0	6.05**	1.51**	.01	-.73
2	2.6	.09	.1	5.1	94.8	106.30**	5.80**	.08	-.60
3	4.4	.10	0	11.1	88.9	120.50**	15.10**	.04	-.56
4	32.7	.48	0	2.0	98.0	3361.90**	69.00**	1.37	-.53
5	49.1	.41	0	5.8	94.2	2489.90**	154.50**	1.29	-.65
6	142.6	1.39	.1	4.0	95.9	3717.00**	158.50**	2.07	-.69
	b_2=+.0093		.05	8.00	91.95				-.63
b. Logarithmic data (coded)									
1	.7160	.39	.4	2.0	97.6	.333641**	.008225**	.001416	-.11
2	1.4082	.48	.2	0	99.8	.874341**	.001188	.001416	-.11
3	1.6405	.45	.2	.4	99.4	.829130**	.004708*	.001416	-.11
4	2.5146	.54	.1	0	99.9	1.123795**	.000777	.001416	+.11
5	2.6911	.51	.2	.3	99.5	.955023**	.004347*	.001416	-.02
6	3.1541	.62	.2	.4	99.4	1.025322**	.005975**	.001416	-.19
	b_2=+.079		.2	.5	99.3			h =	-.07

* obtained by separate analyses of variance of arithmetic values for each C.P. group.

Table $X_3 - 4$

"h" values for F_1's. Arithmetic and logarithmic data/ X_3/

a. Arithmetic

	1	2	3	4	5	6	
1		-.52	-.54	-.80	-.85	-.95	
2			-.33	-.64	-.67	-.82	
3				-.47	-.64	-.83	
4					-.50	-.26	
5						-.61	
6							
c.p."h" values	-.73	-.60	-.56	-.53	-.65	-.69	$\bar{h}$ -.63

b. Logarithmic

	1	2	3	4	5	6	
1		-.15	-.06	-.04	-.08	-.24	
2			-.13	-.10	-.06	-.12	
3				+.01	-.13	-.25	
4					+.57	+.09	
5						-.41	
6							
c.p. "h" values	-.11	-.11	-.11	+.11	-.02	-.19	$\bar{h}$ -.07
disregarding two erratic values	-.11	-.11	-.11	-.01	-.09	-.13	-.09

is again substantiated by the four significant "deviations from regression mean squares," two of which are highly significant.

In summarizing, one can conclude that the model which fits the data most exactly is that of logarithmic gene action with a low order of negative dominance, i.e., h= -.05 to -.10. The data fit this model remarkably well.

X_4 - Number of clusters bearing ripe fruit

On examination of the arithmetic P_1-F_1 Table X_4-1 or "h" Table X_4-4, one can readily see that the F_1 values are generally intermediate with only three cases of heterosis. Two of these F_1 values are not significantly higher than the highest parent.

The c.p.r. trend in arithmetic analysis is increasing, indicating, (1) arithmetic gene action with negative dominance, or (2) gene interaction, presumably logarithmic gene action. This trend, although quite severe, is a bit irregular. When such irregularities occur it is interesting to analyze the data by leaving out first one c.p. group and then another, until the analysis becomes uniform, and then attempt to interpret the erratic behavior of the trouble-causing constant parent. First obvious C.P. group to omit is 1, as C.P. 1 is a different species with probably the most divergent gene complexes among the inbreds. Also C.P. 1 generally has an extreme character expression and it is of interest to see if removal of this extreme parent and its F_1's would shift the logarithmic type of regression trend to that of arithmetic gene action.

When C.P. 1 group is removed the trend smoothes out very nicely and is even more severe than before. Conclusions are that C. P. 1 is causing some

of the irregularities found in the entire analysis, however, it alone is not responsible for the increasing trend indicative of logarithmic gene action.

The alternative hypothesis of arithmetic gene action with negative dominance may be disposed of for several reasons. First and most obvious, the F_1 values (all but 2) do not yield negative dominance values. This can be seen in Table X_4-4. Second, heterosis is in a positive and <u>not</u> negative direction. Third, the highest c.p.r. coefficient in analysis omitting C.P. 1 group is greater than one one, indicating gene interaction rather than negative dominance as cause of the trend. Fourth, the highly variable arithmetic F_1 "h" values become much more uniform with logarithmic transformed data. Thus there is little doubt that the non-additive aspects of this gene behavior are of a gene interaction nature - presumably logarithmic, rather than negative dominance.

Because the arithmetic F_1 values, in general, have greater than mid-parental values, one might suggest that a logical scheme would be arithmetic gene action with positive dominance. However, the c.p.r. trend is directly opposed to this model, as positive dominance would cause a decreasing rather than increasing trend. Nevertheless positive dominance is strongly indicated by the several cases of heterosis in a positive direction and over all "h" value = +.57.

On transforming the data to logs, difficulties at once disappear, and a logical interpretation of the data becomes apparent. With the log data the c.p.r. trend is completely reversed yielding a severe positive dominance trend ($b_2 = -.40$) which is also quite uniform. Thus the obvious fact of positive dominance may be accounted for. Further evidence for logarithmic

gene action is indicated by the smoothing out of the erratic, arithmetic F_1 "h" values by the transformation to logs.

In considering the regression analysis of variance, the arithmetic analysis has four significant "deviations from regression mean squares" as contrasted with only two for log analysis, which is in accord with the other criteria. These two highly significant mean squares in the log data also lend support to the hypothesis of domiance with logarithmic action.

Examination of the components of regression analysis of variance in Table X_4-3 is at first rather disappointing because there is not much difference between the arithmetic and log analyses. This is apparently due to the fact that the high amount of positive dominance tends to lessen the metrical bias effect in the arithmetic data.

Exactly this same situation was found in trait X_2, where, log gene action plus high positive dominance resulted in a relatively small value for component D in arithmetic analysis which was not much different from the logarithmic component D. This was in direct contrast to the situation in average fruit weight, X_3, where the metrical bias was not confounded by dominance effects and was measured directly by the large arithmetic component D.

Another point of interest is that the relatively small component D or small mean square for "deviations from regression" indicates that the F_1 values coincide closely with the regression line. This close fit of F_1's to regression line is characteristic of the purely additive scheme of no-dominance and no-interaction between loci. Apparently, this condition of good fit to regression line is also accomplished by logarithmic gene action plus considerable dominance. However, this pseudo-pure-additive effect is quickly detected by (1) F_1 means are consistently higher than

Table X_4-1

Arithmetic mean P_1 and F_1 values for "Number of Clusters" (X_4) constant parent regression coefficients also listed.

Parents	1	2	3	4	5	6
1	133.5	134.4	129.1	77.5	93.5	81.7
2		66.8	84.0	45.1	39.7	35.6
3			41.7	37.3	32.0	28.1
4				13.4	13.0	8.7
5					8.9	9.7
6						6.1
c.p.r. coefficients	.94	.76	.79	.53	.65	.58
c.p.r. coefficients ommitting C.P. 1 group		1.36	.90	.61	.52	.49

Table X_4-2

Logarithmic mean P_1 and F_1 values for "Number of Clusters" (X_4) constant parent regression coefficients also listed.

Parents	1	2	3	4	5	6
1	2.1255	2.1284	2.1109	1.8893	1.9708	1.9122
2		1.8248	1.9243	1.6542	1.5988	1.5515
3			1.6201	1.5717	1.5052	1.4487
4				1.1271	1.1139	.9395
5					.9494	.9868
6	L.S.D. = $t_{.05}\,\hat{\sigma}_d$ = .1054					.7853
c.p.r. coeffs.	.23	.45	.50	.68	.72	.82
c.p.r. coeffs. omitting C.P. 1 group		.53	.47	.68	.62	.74

Table X_4-3

Summary of statistics used in estimating gene action for trait X_4. Both arithmetic and logarithmic data presented.

C.P.	value	C.P. reg.	Components of Regression A. of V. in % E	D	B	Regression M.S.	Deviations from reg. M.S.	Approx-imately	Ave "h" for C.P.
a. Arithmetic data									
1	133.5	.94	2.9	2.5	94.6	2453.2**	136.0	72.3	+.51
2	66.8	.76	.4	1.1	98.5	6736.0**	102.8**	25.0	+.72
3	41.7	.79	.2	0	99.8	7603.8**	13.9	19.0	+.92
4	13.4	.53	.2	.4	99.4	3023.8**	18.9*	5.3	+.30
5	8.9	.65	.1	.4	99.5	4474.1**	23.3**	5.3	+.64
6	6.1	.58	.1	.3	99.6	3492.7**	14.3*	3.9	+.34
	b_1 = +.0028		.6	.8	98.6				$\bar{h}$ = +.57
b. Logarithmic data									
1	2.1255	.23	3.2	4.3	92.5	.040156**	.003139	.001349	+.78
2	1.8248	.45	.6	0	99.4	.240980**	.000577	.001349	+.89
3	1.6201	.50	.4	0	99.6	.334177**	.000598	.001349	+.99
4	1.1271	.68	.2	0	99.8	.618153**	.001196	.001349	+.52
5	.9494	.72	.2	.9	98.9	.603215**	.006538**	.001349	+.84
6	.7853	.82	.2	1.1	98.7	.640072**	.008397**	.001349	+.62
	b_1 = -.4000		.8	1.1	98.1				$\bar{h}$ = +.77

Table X_4-4

"h" values for F_1's involving both arithmetic and logarithmic data. Trait X_4

a. Arithmetic

	1	2	3	4	5	6	
1		+ 1.03	+ .90	+ .07	+ .36	+ .19	
2			+2.37	+ .19	+ .06	- .03	
3				+ .69	+ .41	+ .24	
4					+ .82	-.29	
5						+1.57	
6							
Ave "h"	+ .51	+ .72	+ .92	+ .30	+ .64	+ .34	$\bar{h}$ + .57

b. Logarithmic

	1	2	3	4	5	6	
1		+ 1.02	+ .94	+ .53	+ .74	+ .68	
2			+1.97	+ .51	+ .48	+ .47	
3				+ .80	+ .66	+ .59	
4					+ .85	-.10	
5						+1.46	
6							
Ave "h"	+ .78	+ .89	+ .99	+ .52	+ .84	+ .62	$\bar{h}$ + .77

additive means, and (2) although F_1 points fit the regression line closely, the c.p.r. coefficients do not all remain at .5, but form severe trends as in the cases of characters X_2 and X_4.

Table X_4-5 gives the means and "h" values for the various generations involved with the segregating populations.

Table X_4-5

Means and "h" values for various generations concerned with segregating populations.

		P_1	P_2	F_1	F_2	B_1	B_2	
Means	Arith	6.07	8.95	9.70	8.66	7.54	8.76	
	log	.7832	.9518	.9868	.9375	.8774	.9425	Ave. "h"
"h"	Arith			+1.57	+1.60	+1.04	+.74	+1.24
	log			+1.46	+1.66	+1.23	+.78	+1.28

This particular cross was one of only three that yielded heterosis among F_1's, and the heterosis of F_1 was maintained at least in part among the segregating populations.

The gene model chosen for this trait in considering the over-all range of character expression is that of logarithmic gene action with considerable positive dominance, i.e., "h" on order of +.75. h=+.75 is chosen for the dominance value rather than h =+1.28 from Table X_4-5, because the former value is obtained over a considerable range of character expression involving six parental values. The latter is obtained with only two parents which are at the low extreme of character expression and not very different.

X_5 - Number of fruits per cluster

F_1 values are generally intermediate in relation to their parents as can be seen from the P_1 - F_1 mean Table X_5-1 or Table X_5-3 of "h" values.

In general, however, the F_1's have values greater than the corresponding mid-parental values and two have as many fruits per cluster as their high parent.

The c.p.r. coefficients form a slight and erratic positive dominance trend yielding $b_2 = -.016$. This decreasing trend automatically eliminates the consideration of a logarithmic model, and there are other aspects of the analysis which support the arithmetic scheme. First, from Table X_5-2 it may be observed that the arithmetic components and mean squares do not suggest any non-additive interaction, i.e., dominance or metrical bias. Second, the "h" values are fairly uniform.

In evaluating the amount of dominance, there are four sources of estimation. First, the c.p.r. trend is irregular and weak indicating slight positive dominance. Second, in the regression analysis of variance there are no "deviations from regression mean squares" which are significantly different from the error. Likewise, the percent of component D is extremely small. Both these factors indicate little or no dominance or gene interaction of a general nature. Third, table of F_1 "h" values yield an average $h = +.49$. Fourth, the segregating populations, together with the F_1 involved in the cross, yield an intermediate average "h" of +.16, as may be seen from Table X_5-4.

Range of estimates of "h" by various techniques extends from (0) through +.16 up to +.49. A middle value in the neighborhood of +.15 to +.20 would probably be best to associate with the arithmetically cumulative gene action for the best estimated gene model.

Table X_5-1

Arithmetic mean P_1 and F_1 values for "number of fruits per cluster" (X_5) Constant parent regression coefficients also listed.

Parents	1	2	3	4	5	6
1	10.5	9.4	9.7	8.7	8.4	7.6
2		5.6	6.0	5.0	5.6	4.7
3			6.2	5.7	6.2	5.2
4				3.9	4.9	3.5
5					5.2	4.2
6						2.6
c.p.r. coefficients	.51	.61	.58	.67	.53	.62

Table X_5-2

Summary of statistics used in estimating gene action for trait X_5. Arithmetic values only.

C.P.	value	C.P. reg.	Components of Regression A. of V. in % E	D	B	Regression M.S.	Devs from Regression M.S.	$\sigma_{\bar{x}}^2$	Ave."h" for C.P.
1	10.5	.51	10.5	0	89.5	2.20**	.19	.23	+ .42
2	5.6	.61	.8	.6	98.6	13.74**	.19	.11	+ .51
3	6.2	.58	1.0	.5	98.5	12.31**	.19	.13	+ .59
4	3.9	.67	.6	0	99.4	14.70**	.06	.09	+ .45
5	5.2	.53	1.0	0	99.0	10.27**	.01	.10	+ .60
6	2.6	.62	.8	0	99.2	9.68**	.03	.08	+ .34
	b_2 = -.016		2.5	.2	97.3				+ .49

Table X_5-3

"h" values for F_1's involving arithmetic data only. Trait X_5

Parents	1	2	3	4	5	6	
1		+.55	+.63	+.45	+.21	+.27	
2			+.33	+.29	+1.00	+.40	
3				+.57	+1.00	+.44	
4					+.54	+.38	
5						+.23	
6							
Ave "h" for C.P.'s	+.42	+.51	+.59	+.45	+.60	+.34	$\bar{h}$ = +.49

Table X_5-4

Means and "h" values for various generations concerned with segregating populations.

	P_1	P_2	F_1	F_2	B_1	B_2	
Means	2.58	5.18	4.20	3.93	3.24	4.78	Ave. "h"
h values			+.23	+.08	+.02	+.38	+.16

X_6 - Average number of locules per fruit

This character is the first case in which the segregating populations yield information of considerable consequence to gene analysis.

Examination of the frequency distributions of the F_2, B_1 and B_2 in

Table 42 points out that a major gene differentiating between few and many locules is segregating. The parents involved in the cross are: P_1 = Devon, average locule number 2.68 (i.e., mode of 3 locules), and P_2 = Matchless, average locule number 6.81 (mode apparently 6 locules).

As can be seen from the F_1's and backcrosses, this major gene exhibits partial dominance toward low number of locules. It may be noted also that the expressivity of each phase of the gene is different. Lower locule numbered types form the most distinct class, and the expression of the gene phase for high locule number is undoubtedly more subject to alteration by environmental effects and segregation of minor modifying genes.

The next step in the analysis is the examination of the c.p.r. analysis. As can be noted from Table X_6-1, the arithmetic means yield an increasing c.p.r. trend, indicative of either log gene action or arithmetic action with negative dominance. Choice of the latter scheme is resolved from the following points. (1) The c.p.r. trend is not essentially changed when data are transformed to logs (Table X_6-2). (2) Component D in arithmetic analysis, for all but the first and six groups, is relatively small, and thus does not warrant consideration of metrical bias. (3) On transforming to logs, the D components are not essentially changed, except for C.P. 1 group, and this one shall be discussed later. Thus conclusions arrived at from both segregating populations and P_1 - F_1 analysis are: (1) a major gene is involved, (2) the joint action of this major gene and other minor genes concerned with this character is, in general, that of arithmetic gene action with considerable negative dominance.

Table X_6-1

Arithmetic mean P_1 and F_1 values for "Average number of locules per fruit" (X_6) Constant parent regression coefficients also listed. Data based on complete sample.

Parents	1	2	3	4	5	6
1	2.00	2.00	2.00	2.00	2.00	2.13
2		2.01	2.01	2.03	2.11	3.17
3			2.00	2.03	2.16	3.23
4				2.09	2.41	3.83
5					2.68	3.65
6						6.81
c.p.r. coefficients	.03	.24	.26	.37	.31	1.03

Table X_6-2

Logarithmic mean P_1 and F_1 values for "Average number of locules per fruit" (X_6) Constant parent regression coefficients also listed.

Parents	1	2	3	4	5	6
1	.2999	.3010	.3010	.3010	.3010	.3290
2		.3021	.3032	.3081	.3243	.5007
3			.3015	.3073	.3349	.5087
4				.3193	.3827	.5827
5					.4283	.5623
6	L.S.D. = $t_{.05}\hat{\sigma}_d$ = .0165					.8328
c.p.r.	.05	.38	.39	.52	.43	.83

Table X_6-3

Summary of statistics used in estimating gene action for trait X_6. Both arithmetic and logarithmic analyses are given.

C.P.	value	C.P.	Components of Regression A. of V. in % E	D	B	Regression M.S.	Dev. from Regression	$\bar{\sigma}_{\bar{x}}$	Ave. "h" for C.P.
a. Arithmetic									
1	2.00	.028	6.1	13.0	8.09	.0100**	.0022*	.0007	-.99
2	2.01	.244	.1	0	99.9	1.0300**	.0000	.0009	-.68
3	2.00	.255	.1	0	99.9	1.1300**	.0000	.0009	-.45
4	2.09	.373	0	.3	99.7	2.4400**	.0070**	.0010	-.40
5	2.68	.310	.1	1.5	98.4	1.7600**	.0270**	.0010	-.54
6	6.81	1.029	.2	55.3	44.5	.3700**	.4600**	.0017	-.55
	b_L = +.17		1.1	11.7	87.2				-.60
b. Logarithmic									
1	.2999	.054	5.6	0	94.4	.000594**	.000011	.000033	
2	.3021	.375	.1	.5	99.4	.029177**	.000170**	.000033	
3	.3015	.390	.1	.1	99.8	.031611**	.000071	.000033	
4	.3193	.521	.1	0	99.9	.057607**	.000039	.000033	
5	.4283	.433	.1	2.1	97.8	.041786**	.000942**	.000033	
6	.8328	.825	.2	55.7	44.1	.008349**	.010545**	.000033	
	b_L = +.95		1.0	9.7	89.3				

Table X_6-4

"h" values for F,'s involving arithmetic data only. Trait X_6.

Parents	1	2	3	4	5	6
1		-1.00	-	- 1.00	- 1.00	-.95
2			+ 1.00	-.50	-.70	-.52
3				- .33	-.53	-.49
4					+ .08	-.26
5						-.53
6						
Ave "h" values dis reg. h = +1.00	-.99	-.68	-.45	-.40	-.54	-.55
		$\bar{h}$ = -.60				

Table X_6-4 estimates the average over-all "h" to be -.60.

"h" value for the proposed gene model may also be estimated from the segregating populations. Table X_6-5 gives "h" estimates for the segregating populations and they approximate the $P_1 - F_1$ h = -.60 rather closely and uniformly.

Table X_6-5

Mean "h" values for various generations concerned with segregating populations. Arithmetic values only.

	P_1	P_2	F_1	F_2	B_1	B_2	
Mean	2.68	6.80	3.65	3.85	3.18	4.96	Av. "h"
"h"			- .53	- .86	-.51	- .79	-.67

Before dropping the gene analysis for this character, one should consider other interesting inferences which are suggested on closer examination of the data. It is of particular interest to examine the influence of the segregation of this major gene on the other variables considered in this thesis. For this study, the distribution of locule number was recorded in tenths of locules so that a frequency for each class was available for F_2, B_1 and B_2.

Totals for each of the other variables were obtained for every locule class. Then the F_2 locule distribution was divided as nearly as possible into a 1:2:1 ratio and the backcrosses were divided into 1:1 ratios. These main categories with their respective locule range are given as follows:

F_2	freq.	range	B_1	freq.	range	B_2	freq.	range
AA	31	2.3-3.1	AA	43	2.1-3.2	Aa	42	3.1-5.1
Aa	62	3.2-4.4	Aa	41	3.3-4.7	aa	41	5.2-6.9
aa	32	4.5-6.0						

Means of the ten different variables for each of these genotypic categories are given in Table X_4-6. Conclusions which may be drawn from this analysis are as follows: (1) Segregation of the locule gene, apparently, does not affect (a) flowering date and (b) yield. (2) Segregation of the locule gene affects rather strongly (a) average fruit weight, and (b) total number of fruit per plant. (3) Other characters affected to some extent are X_4, X_5, X_7, X_9, and X_{10}. A consistent trend in all segregating populations is an essential criterion for determining if the different locule gene phases have differential effects.

These correlated responses to the segregation of this major gene may be due to: (1) multiple effects of the single gene, (2) close linkage of

this locule gene with genes which have large effects on other variables, and (3) combination of one and two.

Before discussing these possibilities, it is necessary to review previous studies bearing on this problem. Lindstrom (16) discovered a major size gene on the first chromosome closely linked with the qualitative genes Dd (dward) and Pp (Peach). Yeager (32) reported on a major locule gene, denoted Lclc (probably the same locus as involved in this study), and found that it was located also on the first chromosome but at the opposite end of the chromosome from the Dd locus. In fact, he obtained 46.9 percent crossing over between these two loci (Dd and Lclc). Both McCleskey (20) and Yeager (32) have found highly significant correlations between locule number and fruit weight.

McCleskey (20) with use of partial correlations, found that, "the apparent high positive correlation between locule number and weight (of fruit) is due to each being highly correlated with equatorial diameter". It would seem to this writer that measurements of equatorial diameter and weight of fruit are essentially different scales of measurement for the same character - that of fruit size. Regardless of the interpretation there is apparently a direct or indirect association between locule number and fruit size.

The genetic interpretation of this association is of interest. Since two major genes, one for locule number and one for fruit weight, are probably located on the same chromosome there is undoubtedly some linkage effect. However, these genes are estimated to be quite far apart on the chromosome so that the linkage effect may be small. Looking ahead to the section on genetic correlations, it may be noted that there occurs a

Table X_6-6

Influence of the major locule gene on other variables as found in F_2, B_1 and B_2 generations.

Trait		Different phases of major locule gene AA	Aa	Aa*
X_1 Yield	F_2	2224	2015	2033
	B_1		2223	2288
	B_2	2034	2006	
X_2 Total number of fruit per plant	F_2	26.0	30.7	40.0
	B_1		35.7	42.5
	B_2	19.5	24.9	
X_3 Average fruit weight	F_2	85.5	65.6	50.8
	B_1		53.8	62.3
	B_2	104.5	80.4	
X_4 Number of clusters per plant	F_2	7.6	8.4	8.8
	B_1		8.1	8.7
	B_2	6.9	7.5	
X_5 Number of fruits per cluster	F_2	3.4	3.7	4.6
	B_1		4.4	4.9
	B_2	2.8	3.3	
X_7 Weight per locule	F_2	17.0	17.3	18.8
	B_1		17.3	19.6
	B_2	17.7	20.1	
X_8 Flowering date	F_2	17.1	18.0	17.2
	B_1		15.6	15.0
	B_2	16.9	17.3	
X_9 Maturity time	F_2	43.2	42.5	42.6
	B_1		42.1	41.4
	B_2	43.7	41.9	
X_{10} Number of flowers per cluster	F_2	5.1	5.6	6.2
	B_1		6.0	6.7
	B_2	4.7	4.9	

* Symbols used by Yeager (32) are A = lc
a = Lc

genetic correlation of + .74 between these two traits. Since the derivation of this genetic correlation is thought to be nearly freed from linkage effects, it is suggested that the major locule gene, itself, does influence directly or indirectly size of fruit. There is the possibility, also, that Devon and Matchless do not have allelic differences at the locus for the major size gene because Devon is a relatively large fruited variety for such a low numbered locule type. In that case, linkage involving these genes would not be involved.

It is possible to integrate action of the locule gene with other characters, by merely observing the trends, in Table X_6-6, which are associated with different phases of the major locule gene. An increase in number of locules (AA → Aa → aa) is associated with a decided increase in fruit weight and with fewer fruits per plant, and fewer flowers and fruits per cluster. Number of clusters bearing ripe fruit follows the same general scheme. The fact that weight per locule decreases with increased fruit weight is due to increased locule number.

Thus it can be seen that number of locules is correlated with most of the other variables. A gene with large effect on locule number, then, influences these correlated characters either directly or indirectly and to varying degrees. In this way it may be pointed out that this single gene has multiple effects (affects many characters), and actually the only reason it is considered a "locule" gene is that its phases are most clearly observed in this essentially discontinuous character.

One of the most interesting aspects in the study of this character lies in the comparisons of the first three parents. Phenotypically, parents 1, 2, and 3 are all two-loculed, but genotypically they are

obviously different. The relative strengths of the two-locule types may be tested by crossing each with parents 5 and 6 and contrasting the values of the corresponding F_1's. These values taken from Table X_6-1 are as follows:

Parent	Parent value	Crossed by parent 5	Crossed by parent 6
1	2.00	2.00	2.13
2	2.01	2.11	3.17
3	2.00	2.16	3.23

It can be seen that there are genotypic differences among parents 1, 2 and 3, although they are phenotypically the same. The differences are mostly between parent 1 and the other two parents. The question then arises - what are the causes of these genetic differences? At least two possibilities may be suggested. First, different strength of major alleles may be involved. It has previously been pointed out that a major gene locus for locule number exists. Two alleles have been discussed so far; allele a associated with large number of locules in the Matchless parent; allele A found in Devon, having a mode of 3 locules. Carrying the allelic series concept further, the two-locule parents may be associated with a different allele A_2. Moreover, the alleles found in parents 1, 2 and 3 may each differ in strength, undetected phenotypically until crosses are made with different genetic backgrounds. This concept has been discussed by various authors probably most thoroughly by Harland (10) and most recently by Stern (29) who coined the term " iso-allele ". Stern (29) states that, "The existence of iso-alleles, defined as alleles indistinguishable except by special tests, is probably a general phenomenon".

The second possibility is that the same major allele A_2 may be present

in all three parents 1, 2 and 3, but it is combined with different modifier complexes in each of these parents. In other words, $\underline{A_2}$ would be combined with a fairly powerful modifier complex in parent 1 to give a stronger two-locule genotype than parents 2 and 3.

Information is available which may suggest the correct hypothesis. Modifier action would imply gene inter-action, which would yield metrical bias in an arithmetical scale, but which would be removed by log transformation. From Table X_4-3, component D of the arithmetic analysis for C.P. 1 has a relatively high value (13.0%), which is entirely removed by log transformation. This would suggest that the excessive strength of C.P. 1 for the two locule expression is due to an accumulated modifier complex which bolsters the action of the major two-locule allele.

$\underline{X_7}$ - Weight per locule

Arithmetic F_1 values for this character are all intermediate between the parents except two, as may be seen in Table X_7-1 or the "h" Table X_7-4. In general, the F_1 means are below the midparental values giving an over-all "h" value of -.43.

The c.p.r. trend with arithmetic data is uniformly increasing with increasing c.p. values. This suggests for a model, either, (1) arithmetic gene action with negative dominance, or (2) gene interaction - presumably of logarithmic nature.

Log transformation drastically reduces the severe arithmetic c.p.r. trend to a no-trend scheme in which all c.p.r.'s are essentially .5, as may be noted in Table X_7-2. This change brought about in trends furnishes evidence that the arithmetic trend is due to logarithmic gene interaction -

removable by transformation. Examination of regression analyses of variances in Table X_1-3 offers further evidence in favor of the log model. With arithmetic analysis, component (D) is relatively small; yet on transformation it almost entirely disappears. Probably of more significance is the fact that even though component D in the arithmetic analysis is small, five out of six of the "deviations from regression" mean squares are highly significant. On log transformation none of the six is significantly different from the error mean square, indicating that metrical bias effects entered into the arithmetic mean squares. Thus considerable evidence points to the choice of logarithmic scale as basis for the best fitting gene model.

The next consideration is that of degree of dominance. C. p. r.'s of log values are all remarkably uniform and essentially .5, which is indicative of no-dominance scheme. A hypothesis of essentially no-dominance is strongly upheld in regression analysis of variance of log data. The relatively small "deviations from regression" mean square with complete lack of significance indicate that the F_1 values fall closely to their respective c.p.r. lines. This together with the fact that all these regressions have, approximately, values of .5, indicates that there is little non-additive (dominance) interactions. Low values of component D also substantiate this.

Examination of F_1 "h" values in Table X_1-4 indicates a low order of negative dominance, which fits other dominance estimates quite well, (except for three erratic values, 4x5, 4x6, and 5x6).

Values of "h" for the segregating populations, Table X_1-5 yield high negative estimates which closely correspond to the particular F_1 value concerned, but which are much higher than the over-all F_1 "h". It should

Table X_7-1

Arithmetic mean P_1 and F_1 values for "Weight per locule" (X_7) Constant parent regression coefficients also listed.

Parents	1	2	3	4	5	6
1	.3	.5	.7	1.9	2.1	2.1
2		1.3	1.6	4.0	4.9	4.8
3			2.2	5.9	5.8	5.0
4				15.4	19.0	19.0
5					18.2	18.2
6						20.5
c.p.r.	.09	.21	.24	.83	.83	.93

Table X_7-2

Logarithmic mean P_1 and F_1 values (coded) for X_7. Constant parent regression coefficients and least significant difference also given.

Parents	1	2	3	4	5	6
1	.4150	.7076	.8513	1.2833	1.3222	1.3160
2		1.1072	1.2068	1.6021	1.6902	1.6794
3			1.3404	1.7723	1.7604	1.7024
4				2.1875	2.2792	2.2783
5					2.2601	2.2596
6	L.S.D. = $t_{.05}\,\hat{\sigma}_d$ = .1082					2.3118
c.p.r.	.51	.51	.48	.54	.52	.53

Table X_7-3

Summary of statistics used in estimating gene action for trait X_7. Both arithmetic and logarithmic analyses given.

C.P.	value	C.P. reg	Components of Regression A. of V. in % E	D	B	Regression M.S.	Devs from Regression M.S.	$\hat{\sigma}^2_{\bar{F}}$	Ave "h" for C.P.	
a. Arithmetic										
1	.3	.09	.1	0	99.9	2.47**	.01	.01	-.72	
2	1.3	.21	.2	.8	99.0	15.52**	.16**	.04	-.55	
3	2.2	.24	.3	3.6	96.1	21.66**	.88**	.06	-.52	
4	15.4	.83	.2	.4	99.4	275.78**	1.54**	.37	+.03	
5	18.2	.83	.2	2.0	97.8	238.62**	5.23**	.40	-.27	
6	20.5	.93	.2	.8	99.0	255.07**	2.48**	.37	-.55	
	b_L=	+.04	.2	1.3	98.5				$\bar{h}$ = -.43	
b. Logarithmic (coded)										omitting F,'s 19,20 and 21
1	.4150	.51	.4	0	99.6	.344411**	.000323	.001421	-.06	-.06
2	1.1072	.51	.2	0	99.8	.715553**	.000574	.001421	-.08	-.08
3	1.3404	.48	.2	.2	99.6	.672265**	.002960	.001421	-.11	-.11
4	2.1875	.54	.2	0	99.8	.753880**	.000724	.001421	+.38	-.03
5	2.2601	.51	.3	0	99.7	.658348**	.001672	.001421	+.08	-.03
6	2.3118	.53	.2	.3	99.5	.677512**	.003236	.001421	-.18	-.12
	b_L=	+.002	.2	.1	99.7				$\bar{h}$ = +.005	-.07

Table X_7-4

"h" values for F_1's with both arithmetic and logarithmic data. Trait X_7.

Parents	1	2	3	4	5	6	
a. Arithmetic							
1		-.60	-.58	-.79	-.80	-.82	
2			-.33	-.62	-.57	-.64	
3				-.44	-.55	-.69	
4					+1.57	+.41	
5						-1.0	
6							$\bar{h}$
Ave "h" values for C.P.	-.72	-.55	-.52	+.03	-.27	-.55	-.43
b. Logarithmic							
1		-.15	-.06	-.02	-.02	-.05	
2			-.15	-.08	+.01	-.05	
3				+.02	-.09	-.25	
4					+1.53	+.46	
5						-1.02	
6							$\bar{h}$
Ave. "h" values for C.P.	-.06	-.08	-.11	+.38	+.08	-.18	+.005
Ave. "h" omitting +1.53 +.46 +1.02	-.06	-.08	-.11	-.03	-.03	-.12	-.07

be pointed out that the segregating populations result from a cross of parents with relatively large number of locules per fruit whereas the rest of the F_1's, in general, are two-locule or low-locule types.

Table X_7-5

Means and "h" values for various generations concerned with segregating populations.

		P_1	P_2	F_1	F_2	B_1	B_2	
Means	Arith.	18.1	20.5		17.8	18.8	19.1	
	log	1.2577	1.3118		1.2504	1.2742	1.2810	Av. h
"h"	Arith.			-1.00	-2.50	+.17	-1.33	-1.17
	log			-1.02	-2.54	+.22	-1.27	-1.15

In summary, for the parental types used in this study, and considering the extreme range of character expression, the criteria indicate, in a remarkably consistent fashion, that the most appropriate gene model for this character would be, cumulatively logarithmic gene action with slight negative dominance on the order of $\bar{h} = -.05$ to $-.10$.

X_8 - Flowering date

On examination of Table X_8-1 of arithmetic means or the "h" Table X_8-3, it may be noted that most F_1 values are intermediate between parental means, and, in general, are below (or earlier than) the midparental values.

The c.p.r. trend is irregular but slightly increasing, $b_2 = +.002$ (Table X_8-2) with increasing parental values. Log transformation is not considered appropriate, for the following reasons. (1) The c.p.r. trend

is not pronounced. With logarithmic gene action a severe trend would be expected. (2) If the data were transformed to logs, positive dominance would have to be assumed, and it would be difficult to explain the two heterotic values in a negative dominance direction (earlier flowering). For these reasons, arithmetic gene action with slight negative dominance is assumed.

Amount of dominance to be associated with the gene model is the next question. The over-all "h" value for the F_1's is -.53, however, it should be noted that such a high estimate is due considerably to the two heterotic values of -2.60 and -3.33. If these two values are left out, the average "h" turns out to be -.15, a value much more in line with the regression trend estimate. This points out the danger of using "h" values as sole basis of estimating dominance deviations. If the parents are close together, as they are in both of these cases, then random errors associated with the estimates of parental and F_1 means may cause considerable fluctuation and unreliability to enter into estimates of "h". Then, just one or two major erratic values may considerably alter the average "h" value for the entire table. Theoretically if parents and F_1's could be measured without error, the scheme would be valid for all degrees of difference between the two parents.

It is interesting to point out, particularly with this trait, that the different constant parents may have quite different average dominance tendencies. Parent 1 exhibits strong negative dominance in all combinations. Parent 3 on the other hand produces positive dominance in four out of five of its hybrids. The rest are quite variable.

"h" values resulting from the segregating populations are highly

erratic as may be seen from Table X_8-4. Probably the reason for this is that apparently there is little genetic difference between the parents and most variation occurring is environmental.

Table X_8-1

Arithmetic mean P_1 and F_1 values for "Flowering date" X_8. Constant parent regression coefficients also listed.

Parents	1	2	3	4	5	6
1	5.0	4.6	6.8	10.7	8.4	8.1
2		5.5	9.8	13.5	11.5	11.6
3			11.7	14.7	14.0	14.9
4				20.4	16.4	18.4
5					16.2	15.5
6						16.8
c.p.r. coefficients	.38	.57	.48	.48	.45	.56
c.p.r. disregarding C.P. 1 group		.42	.36	.37	.33	.43

L.S.D. $= t_{.05}\,\hat{\sigma}_d = 2.2$

Choice of gene model for the over-all range of character expression is that of arithmetic gene action with slight negative dominance.

Table X_8-2

Summary of statistics used in estimating gene action for trait X_8. Analysis based on complete sample, and analysis of variance and covariance on arithmetic values

C.P.	value	C.P. reg.	Components of Regression A. of V. in % E	D	B	Regression M.S.	Deviations from reg. M.S.	$\hat{\sigma}_{\bar{F}_1}^2$	Ave. "h" for C.P.	"h" values disregarding 2 heterotic values
1	5.0	.38	3.1	0	96.9	19.27**	.27	.59	- .84	- .40
2	5.5	.57	1.3	0	98.7	44.85**	.40	.59	- .39	+ .17
3	11.7	.48	1.3	2.8	95.9	45.77**	1.93*	.59	- .02	- .02
4	20.4	.48	1.9	3.5	94.6	29.10**	1.64*	.59	- .30	- .30
5	16.2	.45	1.6	3.2	95.2	36.65**	1.81*	.59	- .90	- .29
6	16.8	.56	1.1	3.2	95.7	55.26**	2.43**	.59	-.72	- .06
		b_1 = +.001741	1.7	2.1	96.2				- .53	- .15

Table X_8-3

"h" values for F_1's. Arithmetic values only. Trait X_8.

Parents	1	2	3	4	5	6
1		- 2.60	-.46	- .26	- .39	-.47
2			+.39	-.07	+.12	+.08
3				-.31	+.02	+.25
4					-.90	-.11
5	$\bar{h}$ = -.53					- 3.33
6						
Ave "h"	- .84	-.39	-.02	-.30	-.90	-.72
Ave "h" disregarding 2 heterotic values	-.40	+ .17	-.02	-.30	-.29	- .06

Table X_8- 4

Means and "h" values for various generations concerned with segregating populations. Arithmetic data only.

	P_1	P_2	F_1	F_2	B_1	B_2	
Means	16.1	16.8	15.5	17.6	15.4	17.3	Av."h"
"h"			- 3.33	+ 6.57	- 5.00	+ 3.86	+.53

X_9 - Maturity time (Number of days from first flower to first ripe fruit)

A look at the arithmetic means, Table X_9-1 or at the "h" Table X_9-4, shows that all but two F_1's exceed the earliest parent, i.e., negative heterosis.

The c.p.r. trend is sharply inceasing beginning with -.44 and ending with +1.73. Since the range of coefficients is beyond 0 to +1, this is strongly indicative of a model exhibiting over-dominance. This is the scheme, whereby, a low x low may yield slight heterosis, a high x high may yield slight heterosis, but a low x high will yield maximum heterosis as more loci are involved with over-dominance.

The arithmetic c.p.r. trend, although definitely positive, is slightly irregular. The best way to determine the misbehaving constant parent is to plot the F_1 values for each constant parent group. When this is done, it can be seen that C.P. 4 group is the irregularly behaving one. On removing this group from the analyses the F_1 values fit the regression lines more closely, as may be seen by comparing the percent for component D in the arithmetic analyses in Tables X_9-3 and X_9-5. When all parents are considered, the average percent for D is 28.1 and when C.P. 4 group is removed, this reduces to 11.7%. Even though the "deviations from regression" are reduced with removal of C.P. 4 group, the over-dominance trend not only remains, but smooths out and becomes an almost "perfect" over-dominance trend extending from -.76 to +1.67.

The question now arises, is the gene action arithmetic, or is logarithmic action partly responsible for the increasing trend. Analysis on data transformed to logarithms supplies the answer by considering the

Table X_9-1

Arithmetic P_1 and F_1 means for "Maturity time" (X_9). Constant parent regression coefficients, and least significant difference also listed.

Parents	1	2	3	4	5	6
1	38.3	39.1	37.5	34.1	35.5	35.4
2		40.2	38.1	33.8	35.4	35.3
3			41.1	37.0	39.1	38.3
4				40.7	40.5	42.9
5					42.6	42.5
6	L.S.D.= $t_{05}\ \hat{\sigma}_d$ = 1.4					44.9
c.p.r.	-.44	-.51	+.18	+1.51	+1.10	+1.73
c.p.r. disregarding C.P. 4 group	-.76	-.64	+.15		+1.15	+1.67

Table X_9-2

Logarithmic values for P_1 and F_1 means for "Maturity time" (X_9). Constant parent regression coefficients also listed. (Coded data)

Parents	1	2	3	4	5	6
1	.5831	.5922	.5740	.5328	.5505	.5494
2		.6039	.5814	.5293	.5490	.5477
3			.6133	.5682	.5924	.5831
4				.6097	.6072	.6325
5					.6289	.6284
6						.6521
c.p.r.	-.87	-.72	+.16		+1.22	+1.74
c.p.r. disregarding C.P. 4 group						

Table X_9-3

Summary of statistics used in estimating gene action for trait X_9. Analysis based on complete sample, and analyses of variance and covariance on arithmetic values. C.P. 4 group not included.

C.P.	value	C.P. reg.	Components of Regression A of V in % E	D	B	Regression M.S.	Devs from Regression M.S.	$\hat{\sigma}^2_E$	Ave "h" for C.P.
a. Arithmetic									
1	38.3	-.76	2.4	9.9	87.7	7.39**	1.01**	.20	-1.48
2	40.2	-.64	2.0	6.5	91.5	9.38**	.85*	.20	-3.48
3	41.1	+.15	27.4	24.7	47.9	.55	.38	.20	-3.35
5	42.6	+1.14	.6	5.9	93.5	30.26**	2.08**	.20	-3.02
6	44.9	+1.67	.6	11.4	88.0	27.02**	3.66**	.20	-2.13
			6.6	11.7	81.7				-2.69
	b_2 = +.41								
b. Logarithmic									
1	1.5831	-.87	2.4	9.4	88.2	.001004**	.000131**	.000027	
2	1.6039	-.72	1.9	6.8	91.3	.001295**	.000121**	.000027	
3	1.6133	+.16	27.0	24.0	49.0	.000073	.000049	.000026	
5	1.6289	+1.22	.7	6.7	92.6	.003736**	.000296**	.000026	
6	1.6521	+1.74	.7	12.1	87.2	.003330**	.000385**	.000026	
			6.5	11.8	81.7				
	b_2 = +42.5								

Table X_9-4

"h" values for arithmetic F_1 values. Trait X_9.

Parents	1	2	3	4	5	6
1		-.16	-1.57	-4.50	-2.30	-1.88
2			-5.67	-26.60	-5.00	-3.09
3				-19.50	-3.67	-2.47
4					-1.21	+ .05
5						-1.09
6						
C.P."h" value	-2.08	-8.10	-6.58	-10.35	-2.65	-1.70
C.P."h" without C.P. 4 group	-1.48	-3.48	-3.35		-3.02	-2.13

following points: (1) The upward trend with transformed data (Table X_9-2) is even more pronounced than with the arithmetic data; (2) The components of regression analysis of variance are surprisingly similar to those on the arithmetic basis, even when compared group by group. Thus, if average gene action were logarithmic rather than arithmetic, much of component D in arithmetic analysis should be due to metrical bias and removable on transformation (as found in fruit size). Also, the regression trend should be less severe on log analysis. Since such is not the case, arithmetic action is assumed to be best for the model.

Other criteria which point to acceptance of over-dominance are as follows: (1) When all parents are considered, Table X_9-3, five out of six of the "deviations from regression" mean squares are highly significant

and yield 28.1% for component D. Even when C.P. 4 group is not considered, Table X_9-5, four out of five mean squares are significant and yield 11.7 to 11.8% for component D in both arithmetic and log analyses. These facts point to a high degree of non-additive gene interaction. (2) F_1 "h" values, found in Table X_9-4, although erratic are consistently lower valued than -1, which is in line with negative over-dominance.

The segregating populations yield fairly uniform "h" values essentially equal to the F_1 involved in the cross. This can be seen from Table X_9-6.

Table X_9-6

Means and "h" values for various generations concerned with segregating populations. Arithmetic data only.

	P_1	P_2	F_1	F_2	B_1	B_2	
Mean	42.6	44.8	42.5	43.1	42.2	43.9	Av. h
"h"			- 1.09	- 1.09	- 1.73	- .64	- 1.14

In summary, analysis of the data yields consistent evidence to the acceptance of the gene model involving arithmetic action with over-dominance in negative direction. The data fit this model remarkably well.

Before leaving this trait, it is interesting to investigate the phenomenon of over-dominance further. As pointed out, the best descriptive gene model for this trait "Maturity time" is over-dominance. The question now arises, is this over-dominance effect actually due to (1) the heterozygous phase having increased effect over either homozygous

Table X_q-5

Summary of statistics used in estimating gene action for trait X_q. Analysis based on complete sample, and analyses of variance and covariance on arithmetic values. All parents considered.

Parent	Parent value	C.P. reg.	Components of Regression A. of V. in % E	D	B	Regression M.S.	Devs. from Regression M.S.	$\hat{\sigma}^2_{\bar{F}}$	Ave "h" for C.P.
1	38.3	.44	2.9	59.4	37.7	2.78**	4.26**	.20	- 2.08
2	40.2	.51	1.9	40.4	57.7	6.12**	4.34**	.20	- 8.10
3	41.1	.18	16.8	32.8	50.4	.80*	.59	.20	- 6.58
4	40.7	1.51	.3	3.4	96.3	56.97**	2.19**	.20	-10.35
5	42.6	1.10	.6	10.6	88.8	28.26**	3.55**	.20	- 2.65
6	44.9	1.73	.5	22.2	77.3	29.03**	8.50**	.20	- 1.70
	b_i =	+ .34	3.8	28.1	68.1				$\bar{h}$ = -5.24

condition or (2) the interaction of traits, in each of which gene action is not of true over-dominance nature. An excellent example of the latter effect has already been demonstrated in yield.

Factors which are associated with maturity time are (1) average relative fruit growth rate per day, and (2) average fruit size. Since the second character has been analyzed already, there just remains the analysis of rate of growth.

Relative fruit growth per day is calculated from the final fruit weight and the length of time from the date of flowering to ripe fruit as follows:

$$\text{rate} = \frac{\text{Log}\{\text{fruit weight (decigrams)}\}}{\text{MATURITY TIME}} = \frac{\log W}{T},$$

or expressing it another way $\log W = rt$, or $W = e^{rt}$.

The P_1 and F_1 rates are given in Table X_9-7. To demonstrate how these values were calculated consider the F_1 (4x5). Fruit weight of this F_1 is 45.0 gms. or 450 decigrams, and maturity time is 40.4 days. The relative growth rate per day, then, is log (450)/40.4 = .0655. This value is recorded in Table X_9-7 as 655. The c.p.r. trend with all parents considered is a rather irregularly decreasing trend ($b_2 = -5.9$) indicating positive dominance. When C.P. 1 group is omitted, the c.p.r. trend smoothes out nicely to give a consistently decreasing trend ($b_2 = -11.8$) indicating a fair degree of positive dominance.

Positive dominance is further substantiated by examination of the F_1 "h" Table X_9-8. All F_1's but two show some positive dominance and the average for the table is $\bar{h} = +.56$. There is one erratic value of h = +4.29 which results from parents having very similar values, i.e., 618 and 632. When this "h" is disregarded the table becomes more uniform and has as average of $\bar{h} = +.29$. It is interesting to note again that consistent differences occur between some of the c.p. groups, i.e., C.P. 1 is con-

Table X_9-7

P_1 and F_1 means for "Average relative fruit growth per day". Constant parent regression coefficients also given.

Parents	1	2	3	4	5	6
1	187	258	306	464	457	464
2		350	396	564	569	616
3			399	562	536	577
4				618	655	668
5					632	665
6						702
c.p.r.	.63	.71	.54	.39	.40	.41
c.p.r. disregarding C.P. 1 group		.73	.53	.33	.35	.26

Table X_9-8

F_1 "h" values for "Average relative fruit growth per day".

Parents	1	2	3	4	5	6	
1		-.13	+.12	+.29	+.21	+.08	
2			+.88	+.60	+.55	+.51	
3				.49	+.18	+.17	
4					+4.29	+.19	
5						-.06	
6	Ave. h = +.56						
Ave c.p."h"	+.11	+.48	+.37	+1.17	+1.03	+.18	Ave $\bar{h}$ +.56
Ave "h" disregarding value +4.29	+.11	+.48	+.37	+.39	+.22	+.18	+.29

sistently low $\bar{h}_i = +.11$, and C.P. 2 is consitently high $\bar{h} = +.48$. However, it may be concluded that in general relative growth rate exhibits considerable positive dominance for faster growth rate.

Now it is possible to summarize the facts necessary to make deductions concerning the nature of the over-dominance phenomenon exhibited by character X_9, "Maturity time".

1. Gene action for fruit weight is logarithmic with partial dominance for small size. On arithmetic scale, the F_1 fruits are close to the smallest parent.
2. Growth rate in tomato fruits exhibit dominance for fast growth rate.
3. Growth rates differ among the parents - the larger the fruit the faster the rate.

With these facts in mind it is readily seen how the interaction of the expression of growth rate and size of fruit result in over-dominance for maturity time. When two parents of approximately equal sizes and rates are crossed, the F_1 will be essentially the same as the parents, e.g. small (slow rate) x small (slow rate) or large (fast rate) x large (fast rate). However, when a large fruited parent (fast growth rate) is crossed onto a small fruited parent (slow growth rate), the F_1 will have a relatively fast growth rate and a restricted small size fruit, thus yielding a much shorter maturity time than either parent.

X_{10} - Number of flowers per cluster

Analysis of this character is of interest, mainly, because the best fitting gene model apparently depends on the range of character expression.

Examination of the Table of arithmetic means X_{10}-1 or Table X_{10}-4 of "h" values shows that all but two F_1 values are intermediate between their respective parents. It should be noted from the arithmetic "h" Table X_{10}-4 that all F_1's involving C.P. 1 show strong negative values, whereas only one of the remaining ten F_1's is negative.

The c.p.r. trend with arithmetic data, when all F_1's are considered, is erratic but slightly increasing (b_2=+.0107). That this logarithmic effect is due mostly to C.P. 1 group, may be noted from the following points, (1) with arithmetic values, omitting the C.P. 1 group, the trend is reversed indicating a positive dominance trend (b_2 = -.0445), and (2) when the data are transformed to logs, the c.p.r. trend involving all parents is a fairly uniform partial dominance trend (b_2 = -.3515), and (3) from the arithmetic F_1 "h" Table X_{10}-4, it has already been pointed out that the C.P. 1 group exhibits large (-h) values. When transformed to logs most of the negative effects disappear, indicating presence of loga-rithmic gene action. Thus it is rather evident that C.P. 1 is so divergent in expression that it takes log transformation to bring it and its F_1's into line with the other parents and F_1's.

The trends as well as components of regression analysis of variance (Table X_{10}-3) also point out that for the rest of the parents, disregarding the extremely divergent C.P. 1, arithmetic action with slight positive dominance is the most appropriate model.

Table X_{10}-1

Arithmetic P_1 and F_1 means for "Number of flowers per cluster" X_{10}.

Constant parent regression coefficients given.

Parents	1	2	3	4	5	6
1	19.6	10.6	11.3	10.1	11.3	9.5
2		7.5	7.4	6.3	7.1	6.2
3			7.2	6.5	7.3	6.1
4				4.9	6.5	4.9
5					7.4	6.0
6						4.4
c.p.r	.46	.29	.33	.32	.34	.29
c.p.r disregarding C.P. 1 group		.37	.38	.51	.36	.48

Table X_{10}-2

Logarithmic P_1 and F_1 means for "Number of flowers per cluster" X_{10}.

Constant parent regression coefficients listed.

Parents	1	2	3	4	5	6
1	1.2916	1.0245	1.0539	1.0043	1.0512	.9796
2		.8727	.8663	.7980	.8531	.7889
3			.8549	.8109	.8603	.7832
4				.6893	.8136	.6866
5	L.S.D. = $t_{.05}\hat{\sigma}_d$ = .0300				.8686	.7753
6						.6464
c.p.r.	.26	.37	.41	.48	.41	.48

Table x_{10}-3

Summary of statistics used in estimating gene action for x_{10}. Both arithmetic and logarithmic data presented.

C.P.	Value	C.P. reg	Comps of regression A of V in % E	D	B	Regression M.S.	Deve. from Regression M.S.	$\hat{\sigma}^2_{\bar{E}}$	Ave,"h" for C.P.	Ave h leaving out h = -7.0	c.p.r. without c.p.1 group	Ave h without c.p.1 group and without h=-7.0
A. Arithmetic												
1	19.6	.46	3.5	4.4	92.1	1.940**	.160	.070	-.36	-.36		
2	7.5	.29	.2	.1	99.7	12.780**	.043	.031	-1.38	+.06	.37	+.19
3	7.2	.33	.2	0	99.8	17.160**	.017	.032	+.12	+.12	.38	+.23
4	4.9	.32	.2	.5	99.3	14.600**	.103*	.026	+.29	+.29	.51	+.44
5	7.4	.34	.2	0	99.8	17.690**	.033	.032	-1.40	0	.36	+.12
6	4.4	.29	.2	.3	99.5	11.860**	.063	.023	+.22	+.22	.48	+.36
	b_1=+.01		.8	.9	98.3				-.42	.06	b_1=-.04hm+.27	
B. Logarithmic										Ave."h"[1]		
1	1.2916	.26	3.2	4.2	92.6	.003237**	.000250	.000109		-.09		
2	.8727	.37	.3	0	99.7	.035630**	.000076	.000109		+.11		
3	.8549	.41	.2	.1	99.7	.044416**	.000139	.000109		+.15		
4	.6893	.48	.2	.1	99.7	.051878**	.000152	.000109		+.39		
5	.8686	.41	.2	.3	99.5	.044636**	.000241	.000109		+.05		
6	.6464	.48	.2	0	99.8	.045863**	.000077	.000109		+.33		
	b_1=-.35		.7	.8	98.5					hm=+.16		

[1]Log "h" are calculated without the one erratic h=-8.56

Table X_{10}-4

F_1 "h" values for X_{10}. Both arithmetic and logarithmic data.

Parents	1	2	3	4	5	6	
a. Arithmetic							
1		-.49	-.34	-.29	-.36	-.33	
2			+.33	+.08	-7.0	+.16	
3				+.39	0	+.21	
4					+.28	+1.00	
5						+.07	
6							
C.P."h" value	-.36	-1.38	+.12	+.29	-1.40	+.22	$\bar{h}$= -.42
(Leave out h=-7.0)	-.36	+.06	+.12	+.29	0	+.22	$\bar{h}$=+.06
(Leave out C.P.1 and h=-7.0)		+.19	+.23	+.44	+.12	+.36	$\bar{h}$ = +.27
b. Logarithmic							
1		-.28	-.09	+.05	-.14	+.03	
2			+.28	+.19	-8.56	+.26	
3				+.47	-.21	+.31	
4					+.39	+.87	
5						+.16	
6							
(h for C.P. not including h=-8.56)		-.09	+.11	+.15	+.39	+.05	+.33 $\bar{h}$=+.16

Values of "h" from segregating population are negative as may be seen in Table X_{10}-5.

Table X_{10}-5

Means and "h" values for various generations involved with the segregating generations. Arithmetic values only.

	P_1	P_2	F_1	F_2	B_1	B_2	
Means	4.40	7.40	6.00	5.67	4.84	6.45	Av. h
h			+ .07	-.32	-.45	-.27	-.24

To summarize then, for the range of expression from 4.4 to 7.5 among the parents, the best gene model is arithmetic gene action with slight positive dominance. The best estimate of dominance is from data disregarding C.P. 1 group and the one extremely erratic value h = -7.0. This yields $\bar{h}$ =+.27. The erratic h = -7.0 is ignored because it results from parents which are close together (i.e., 7.4 and 7.5), thus casting considerable doubt as to its reliability. When the entire range of character expression from 4.4 to 19.6 is considered, then the best model is, apparently, logarithmic gene action with positive dominance on order of $\bar{h}$ = .16.

Relationships among the Variables

All statistics involved in determining relationships among variables and making selection indices in this study are estimates which may be associated only with this material of material of equal range of character

expression and measured in the same scale. Heritability, for example, is essentially the proportion of total variance which is assignable to differences between genotypes. It is obvious that heritability values will vary considerably in relation to the range of genotypes under consideration. Scale of measurement may also effect heritability values. Genetic correlations are also influenced by the range of expression. Thus it is quite evident that in plant breeding problems these statistics must be estimated specifically for the material actually to be used, and for the conditions under which selection will operate, and with the appropriate scale of measurement. To sum up, a statement such as, "Heritability for yield in tomatoes was found to be .50", has essentially no meaning without proper qualifications.

The estimates determined in this study are estimates based on extremely divergent genotypes measured in all but two cases on log scale. They are, therefore, much higher than estimates which would be encountered in customary plant breeding material. Table 24 gives components of analysis of variances in percentage for the various variables. Component G is considered as an estimate of heritability.

Relationships among these variables will be discussed in units of three variables. Then all will be combined in a summary form.

Relationships among X_1 (yield), X_2 (number of fruit per plant) and X_3 (average fruit weight).

Table 25 gives the mean square and cross products matrix for variables X_1, X_2 and X_3. Phenotypic, genotypic, and environmental correlations, together with phenotypic and genotypic path coefficients are given in Table

Table 24

Components of analysis of variances in percentage for variables studied.

Trait	Components of analysis of variance in percentage G	R	I	E
a. Log analysis				
X_1 (Yield)	71.5	3.3	12.2	13.0
X_2 (No. of fruit)	97.3	0	1.3	1.4
X_3 (Fruit weight)	97.7	.2	1.9	.2
X_4 (No. of clusters)	91.6	.3	2.2	5.9
X_5 (Fruits per cluster)	60.3	2.0	10.2	27.5
X_6 (No. of locules)	97.3	.2	0	2.5
X_7 (Wt. per locule)	96.8	.2	2.7	.3
X_{10} (No. flowers per cluster)	89.6	.6	2.2	7.6
b. Arithmetic analysis				
X_8 (Flowering date)	81.5	16.8	1.2	.5
X_9 (Maturity time)	71.4	4.6	6.5	17.5

26. Phenotypic correlations were based on mean squares and crossproducts between varieties. Genotypic and so called "environmental" correlations were calculated from components of analysis of variance and covariance as indicated earlier.

Interpretation of genetic correlations is of particular interest. Correlation $r_{G_{23}} = -.970$ means that the genotypic complexes for X_2 and X_3 are highly negatively correlated, i.e., a genotypic increase in fruit

Table 25

Matrix of mean squares and cross-products for the various sources of variation involving variables X_1 (Yield), X_2 (Number of fruit per plant), and X_3 (Average fruit weight).

		X_1	X_2	X_3
X_1	Reps	+.09951**	+.04338*	+.06823**
	Var	+.51170**	-.75352**	+1.28636**
	RxV	+.01894**	+.01693**	+.00254**
	Error	+.00498	+.00435	+.00064
X_2	Reps		+.02839	+.02126**
	Var		+5.66662**	-6.34637**
	RxV		+.01682**	+.00075**
	Error		+.00446	-.00010
X_3	Reps			+.07654*
	Var			+7.60402**
	RxV			+.02549**
	Error			+.00075

weight is associated closely with a genotypic decrease in number of fruits per plant. Correlation $r_{G_{12}} = -.462$ means that among the genotypes as a whole, X_1 (Yield) genotypic values decrease with increase of genotypic values for number of fruits.

At first this result is somewhat startling but it must be remembered

that it is a simple correlation between these two variables and the effect of fruit size is not considered. However, when fruit size is taken into account by considering the partial correlation $r_{G_{12.3}} = +1.0$ (remembering that in a closed circuit such as this, partial correlations should have a value of one, and correct sign (plus or minus) indicated by path coefficient), then it is obvious that in groups of genotypes having exactly the same fruit size, genotypic increase in yield is perfectly associated with genotypic increase in number of fruits.

Table 26

Simple correlations and path coefficients for variables X_1, X_2 and X_3.

	X_1X_2	X_1X_3	X_2X_3
Phenotypic ($r_{P_{ij}}$)	−.443	+.652	−.967
Genotypic ($r_{G_{ij}}$)	−.462	+.664	−.970
Environmental ($r_{E_{ij}}$)	+.923	+.332	−.055
Path coefficients			
Phenotypic	$\frac{b'\ 12.3}{+\ 2.880}$	$\frac{b'\ 13.2}{+\ 3.436}$	
Genotypic	+ 3.091	+ 3.662	

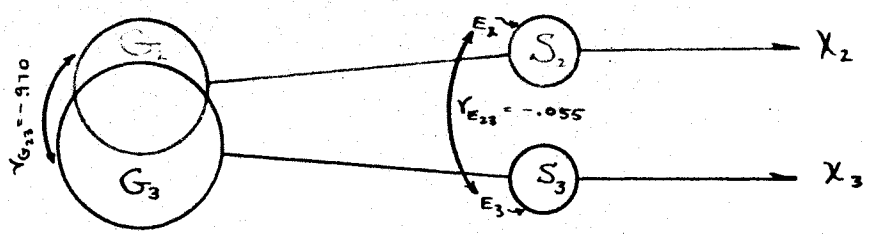
Correlation of $r_{G_{13}} = +.66$ means that genotypic increase in yield is positively associated with genotypic increase in fruit size. This correlation is much less influenced by X_2 than $r_{G_{12}}$ is influenced by X_3, indicating that X_3 is the most influential variate affecting X_1.

Interpretation of phenotypic correlations is essentially the same as for genotypic relationships. Phenotypic correlations are all slightly

smaller than genotypic correlations.

Environmental correlations have interesting interpretations. Environmental factors peculiar to any one plant which increase number of fruit are strongly associated with factors increasing yield, i.e., $r_{E_{12}} = +.923$. The same relationship, although not so strong, holds for yield and fruit size, i.e., $r_{E_{13}} = +.332$. However, environmental factors affecting number of fruits are independent, apparently, of environmental factors affecting size of fruit, i.e., $r_{E_{23}} = -.055$.

The genotypic and environmental relationships for X_2 and X_3 which determine X_1 may be presented as follows:



This would indicate that, although the two sets of genes probably have many genes in common, they operate on subtrates which are relatively independent.

Relationships among X_2 (number of fruits), X_4 (number of clusters) and X_5 (number of fruits per cluster)

Mean squares and cross products are found in Table 27 and correlations and path coefficients in Table 28.

All genetic correlations are extremely high and positive. Of particular interest is the high genetic correlation $r_{G_{45}} = +.943$. This would indicate that a genotypic increase in X_4 is closely associated with a genotypic

Table 27

Matrix of mean squares and cross products for the various sources of variation involving variables X_2 (Total number of fruit per plant), X_4 (number of clusters bearing ripe fruit), and X_5(Number of fruits per cluster).

		X_2	X_4	X_5
X_2	Reps	+ .02839	+ .01716	+ .01105
	Vars	+5.66662**	+4.25679**	+1.40801**
	RxV	+ .01682**	+ .01053**	+ .00606**
	Error	+ .00446	+ .00329	+ .00117
X_4	Reps		+ .06370*	- .04411*
	Vars		+ 3.23833**	+ 1.01657**
	RxV		+ .02428**	- .01383**
	Error		+ .01160	- .00820
X_5	Reps			+ .05274*
	Vars			+ .39170**
	RxV			+ .01990**
	Error			+ .00943

increase in X_5. However, the environmental correlation is highly negative $\gamma_{\epsilon_{45}} = -.784$. This means that environmental effects, peculiar to a particular plant, which increase number of clusters per plant are associated with environmental effects which decrease number of fruits per cluster.

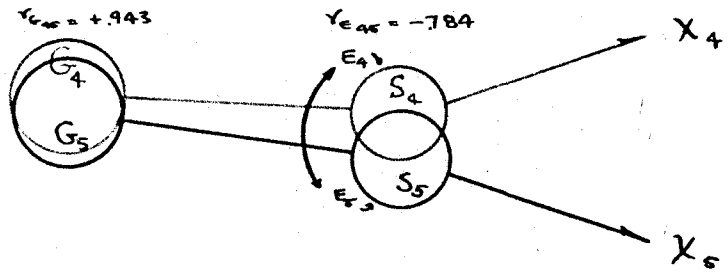
Thus effects of genes apparently are diametrically opposed to random environmental effects. The two large correlations in opposite sign indicate that, (1) the two gene sets have many genes in common, and (2) the two gene sets probably operate on and compete for a common limited substrate. Common substrate is indicated by the largeness of the environmental correlation, and the limited condition is suggested by the evident competition existing causing the negative correlation.

Table 28

Simple correlations and path coefficients for variables X_2, X_4 and X_5.

	X_2X_4	X_2X_5	X_4X_5
Simple correlations			
Phenotypic ($r_{P_{ij}}$)	+.994	+.945	+.903
Genotypic ($r_{G_{ij}}$)	+.996	+.967	+.943
Environmental ($r_{E_{ij}}$)	+.457	+.180	-.784
Path coefficients	b'24.5	b'25.4	
Phenotypic	+.759	+.260	
Genotypic	+.759	+.251	

Thus it is possible to diagram the conditions as follows:



The regression relationships based on phenotypes are essentially similar to those based on genotypic variation.

Relationships among X_3 (Average fruit weight) X_6 (Number of locules per fruit), and X_7 (Weight per locule).

Mean squares and cross products are listed in Table 29 and correlations and path coefficients in Table 30.

The relationships follow the same pattern as those in the second unit (variables X_2, X_4 and X_5), except that the genetic correlations are not so uniformly high.

Of particular interest are the genotypic and environmental correlations between X_6 and X_7. The fairly large correlation $r_{G_{67}} = +.617$ means that the gene sets are fairly similar and that an addition of a gene increasing genotypic expression of X_6 on the average will also increase X_7. In other words the two gene complexes are correlated in positive direction. The fairly large negative environmental correlation, $r_{E_{67}} = -.552$ probably indicates competition for a common limited substrate.

Phenotypic relationships are essentially the same as genotypic and give exactly the same percentages for sources of variation in X_3. Diagramatic relationships are as follows:

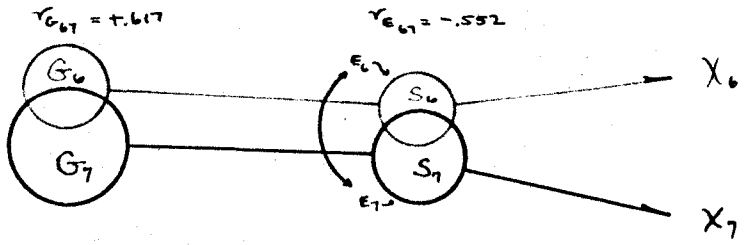


Table 29

Matrix of mean squares and cross products for the various sources of variation involving variables, X_3 (Average fruit weight), X_6 (Number of locules per fruit), and X_7 (Weight per locule).

		X_3	X_6	X_7
X_3	Reps	+ .07654*	+ .00107**	+ .07512*
	Vars	+7.60402**	+1.21896**	+6.37413**
	RxV	+ .02549**	+ .00022**	+ .02525**
	Error	+ .00075	+ .00011	+ .00064
X_6	Reps		+ .00157*	- .00050
	Vars		+ .35658**	+ .86155**
	RxV		+.00055	- .00034
	Error		+.00051	- .00040
X_7	Reps			+ .07529*
	Vars			+5.50252**
	RxV			+ .02558**
	Error			+ .00103

Table 30

Simple correlations and path coefficients for variables X_3, X_6, and X_7.

	X_3X_6	X_3X_7	X_6X_7
Simple correlations			
Phenotypic ($r_{p_{ij}}$)	+.740	+.985	+.615
Genotypic ($r_{G_{ij}}$)	+.742	+.985	+.617
Environmental ($r_{E_{ij}}$)	+.178	+.728	−.552
Path coefficients	b'36.7	b'37.6	
Phenotypic	+.216	+.853	
Genotypic	+.216	+.852	

Over-all genotypic relationships

All variables may be combined and genotypic relationships represented diagramatically in figure 3. The three units are combined illustrating the correlation and path coefficient genotypic relationships existing among the components. Since the phenotypic values are essentially similar, a separate phenotypic diagram is not included. It must be remembered in interpreting these relationships that extreme genotypic ranges for all traits are found in this study, and that logarithms were used as the scale of measurement throughout. These two facts account, largely, for the high heritability and correlation values.

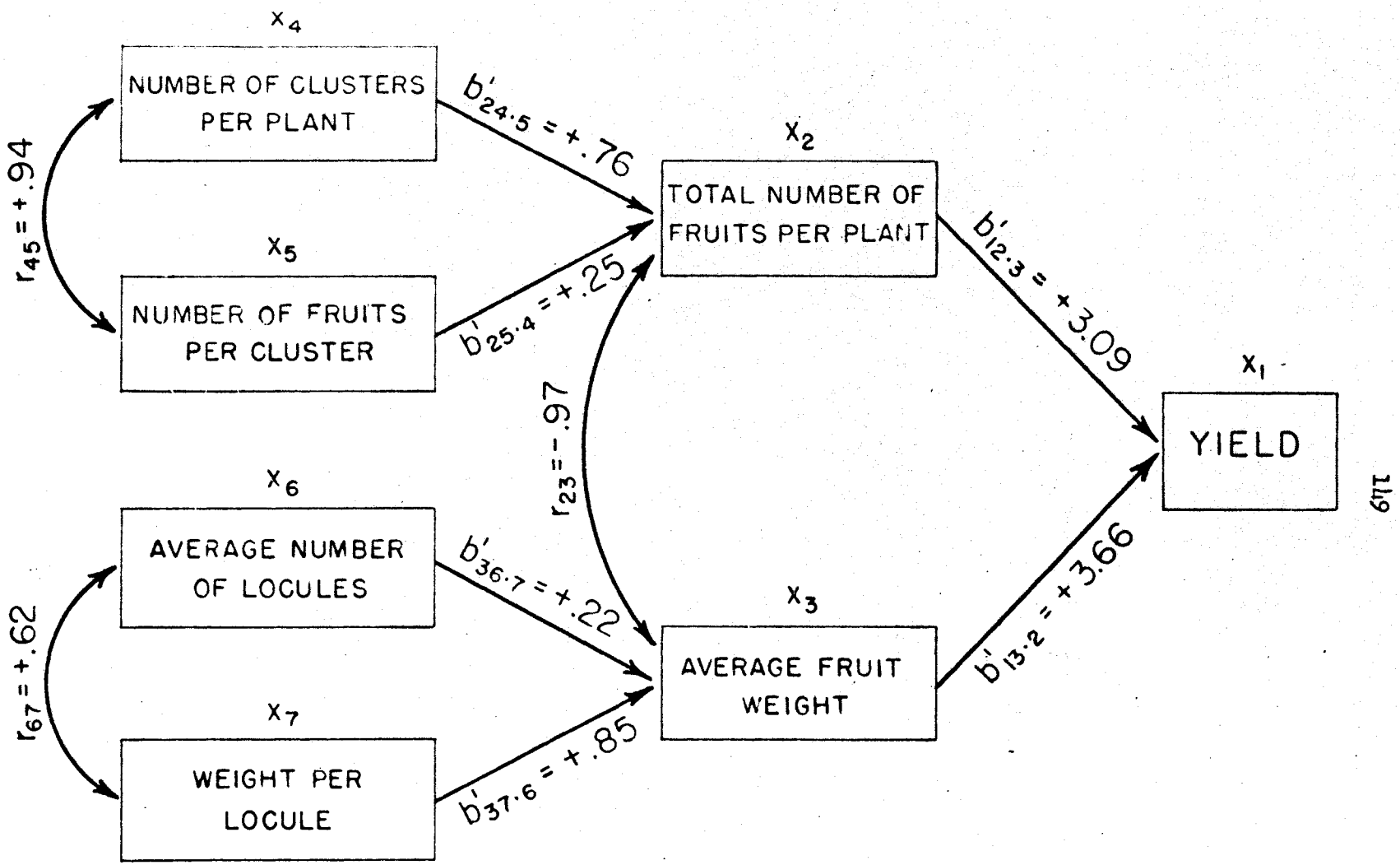


FIGURE 3. DIAGRAMATIC PRESENTATION OF GENOTYPIC RELATIONSHIPS AMONG COMPONENTS OF YIELD IN TERMS OF PATH COEFFICIENTS AND CORRELATIONS

Discriminant function selection index

A selection index could be constructed for any set of variables in this study. However, of particular interest is the combination of yield components to give a selection index for yield itself. The basic idea is that yield has relatively low heritability, generally speaking, in most plant breeding problems, so that selection on yield itself in F_2 plants would involve not only genetic differences but environmental differences as well, which obviously are not inherited and hence their inclusion reduces the effectiveness of selection. Selection for components by use of a discriminant function would tend to minimize this environmental effect. Therefore a selection index is devised for components X_4, X_5, X_6, and X_7.

Components in per cent for analyses of variance and covariance for these variables is given in Table 31. G component is estimate of heritability. For the construction of a selection index, a phenotypic variance-covariance matrix is needed and this is inverted to give the $\|c_{ij}\|$ matrix i.e. $\|\epsilon_{ij}\|^{-1} = \|c_{ij}\|$. (See Tables 32 and 33). Genotypic variance and covariance matrix is also needed, as well as values for a_i's. These values are found in Table 34.

In Smith's (27) presentation, the a_i's are economic values assigned to the variables. In this study, the a_i's are all given equal weight, i.e., $a_4 = a_5 = a_6 = a_7 = 1$. The discriminant function turns out to be:

$$Y = .912 X_4 + .834 X_5 + .884 X_6 + .911 X_7$$

and when the coefficients are adjusted so that they add to four, the function becomes:

$$Y' = 1.031 X_4 + .941 X_5 + .997 X_6 + 1.031 X_7$$

This then may be compared directly to the non-discriminant function based solely on a_i's:

$$Y = 1.000\,X_4 + 1.000\,X_5 + 1.000\,X_6 + 1.000\,X_7$$

Trait X_5 has been discounted most and an examination of Table 31 will largely explain the reason. Heritability for X_5 itself is by far the lowest of the variables and its covariance with X_4 is also lower than other covariances. In general, however, differences are small, but that is what would be expected with this type of data where genotypic ranges are extreme in all traits and heritability values are, for the most part, exceptionally high even in covariance analysis.

This discriminant function should not be used in tomato breeding work, unless verified to be essentially similar to a function developed on the material to be used. The purpose in presenting this section is to demonstrate the methodology in constructing such an index with the type of analysis found herein.

Relationships of "u" gene to variables

The "u" gene, which causes uniformly green skin coloring in developing fruit, was included in the segregating population study in an endeavor to detect linkage relationships between this gene of large effect and genes of the various quantitative variables. The marker gene was brought into the cross by Devon, and, of course, segregated in the F_2 and backcross to Devon.

Table 35 gives the arithmetic means for both the wild phenotype (designated as U) and the recessive (designated u) for parents, F_2 and backcross. Difference between the means of wild and recessive phenotypes

Table 31

Components in percent for analyses of variances and covariances for variables X_4, X_5, X_6 and X_7.

		X_4	X_5	X_6	X_7
X_4	G	97.6	84.4	99.8	98.4
	R	.3	.7	.2	1.5
	I	2.2	2.8	0	.1
	E	5.9	12.1	0	0
X_5	G		60.3	97.7	96.6
	R		2.0	2.1	3.2
	I		10.2	.1	.2
	E		27.5	.1	0
X_6	G			97.3	99.0
	R			.2	0
	I			0	.1
	E			2.5	.9
X_7	G				96.8
	R				.2
	I				2.7
	E				.3

Table 32

Phenotypic variance-covariance matrix ($\|t_{ij}\|$) for variables X_4, X_5, X_6, and X_7.

	X_4	X_5	X_6	X_7
X_4	+3.23833	+1.01657	-.79657	-3.98130
X_5		+.39170	-.28100	-1.27758
X_6			+.35658	+.86155
X_7				+5.50252

Table 33

Inverse phenotypic variance-covariance matrix for variables X_4, X_5, X_6, and X_7. $\|c_{ij}\|$

	C_4	C_5	C_6	C_7
C_4	+4.77117	-2.10730	+2.95811	+2.49970
C_5		+16.19792	+4.26942	+1.56764
C_6			+8.38145	+1.81928
C_7				+2.06951

Table 34

Genotypic variance-covariance matrix for variables X_4, X_5, X_6, and X_7. Values are actually $18\hat{\sigma}^2_{G_{ij}}$. a_i's are also listed.

	X_4	X_5	X_6	X_7
X_4	+3.21405	+1.03040	-.79686	-3.97907
X_5		+.37180	-.28145	-1.27976
X_6			+.35603	+.86192
X_7				+5.47694
a_i's	+1.0	+1.0	+1.0	+1.0

(designated as d = u - U) is also given. Criterion for establishment of a relationship, is that "d" for both F_2 and backcross must be consistently positive or negative and relatively large.

It would appear that the "u" gene, either because of multiple effects or linkage, is associated with greater locule number, greater fruit size, greater number of fruits per cluster, and also, apparently, causes fruits to take longer to mature. However, direct linkage relationships are questionable because the association as found in the F_2 and backcross are directly opposite to the association as found in the parents in all but one trait X_5. In this case Devon (having the recessive "u" gene) has larger number of fruits per cluster than Matchless (having the wild allele) and in the F_2 and backcross the "u" recessives average greater number of fruits per cluster than the wild type. However, in the other three characters X_3, X_4, and X_9, this association is directly opposite to the results which would be expected with linkage.

An explanation might be somewhat as follows. Apparently, there is no difference between "U" and "u" types for number of flowers, but quite a large difference between number of fruits that set and ripen. This would mean that the "u" gene is associated with good fruit set indicating that the "u" gene causes directly or indirectly better food competitive power for fruit embryos. This same power would lead to longer maturity time, larger fruits and greater number of locules. Such explanation is mere speculation. However, it is difficult to suppose that the relationships found are merely accidental.

Table 35

Relationships of "u" gene with other variables as found in F_2 and backcross populations.

Trait		U(normal)	u (recessive)	d= u-U
X_1 Yield	P	1996	2253	+257
	F_2	2005	2392	+387
	BC	2246	2265	+ 19
X_2 Total number of fruits per plant	P	15.9	40.6	+24.7
	F_2	30.9	34.1	+ 3.2
	BC	39.7	38.4	- 1.3
X_3 Average fruit weight	P	142.6	49.1	-93.5
	F_2	64.9	70.2	+ 5.3
	BC	56.6	59.0	+ 2.4
X_4 Number of clusters	P	6.1	9.0	+ 2.9
	F_2	8.2	8.8	+ .6
	BC	9.2	7.7	- 1.5
X_5 Number of fruits per cluster	P	2.6	5.2	+ 2.6
	F_2	3.8	3.9	+ .1
	BC	4.3	5.0	+ .7
X_6 Number of locules per fruit	P	6.8	2.7	- 4.1
	F_2	3.7	4.0	+ .3
	BC	3.0	3.2	+ .2
X_7 Weight per locule	P	20.5	18.1	- 2.4
	F_2	17.5	17.8	+ .3
	BC	18.7	18.2	- .5
X_8 Flowering date	P	16.8	16.2	- .6
	F_2	17.6	17.5	- .1
	BC	15.3	15.3	0
X_9 Maturity time	P	44.9	42.6	- 2.3
	F_2	42.6	43.4	+ .8
	BC	40.9	42.4	+ 1.5
X_{10} Number of flowers per cluster	P	4.4	7.4	+ 3.0
	F_2	5.6	5.6	0
	BC	6.3	6.5	+ .2

DISCUSSION

The problem of genetic analysis of quantitative characters (which includes the phenomenon of heterosis) is perhaps one of the most complex, difficult, and yet important problems of theoretical and applied genetics. The general approach to the problem has been to examine these traits in a measurably quantitative manner using various statistical techniques. The first step is to demonstrate that these characters are under direct control of genes. Such heritable control has long been established for many traits and now assumed to be universally true. The second phase consists of setting up models of hypothetical gene action to describe average gene effects, and estimating genotypic relationships among the various traits considered. In other words, the second phase consists of getting a closer perspective of the genetic behavior involved, generally through a statistical approach. The third phase would involve individual gene analysis and a physiological interpretation of single gene effects. At the present time such a study is limited to the occasional genes exhibiting large effects which are known to exist. The major locule number alleles segregating in this study would be an excellent example. However, new techniques must be developed in order to study the majority of genes having smaller effects. An excellent example of such single gene analysis is the more recent study with sex determining genes in which at least seven individual gene pairs have been isolated for this quantitative character. (See Gowen (7)).

The tomato study involved in this thesis is mainly concerned with the second phase, that of getting fully acquainted with the average genetic behavior of the traits and their relationships among each other. However,

the study does lead into phase three with the breakdown of complex genetic systems into simpler parts and also the analysis of a major locule gene and demonstrating the various correlated responses to its segregation.

More concisely, the material presented in the foregoing sections demonstrate how a complex character may be resolved into components, and how analysis of non-segregating populations (P_1's and F_1's) may lead to estimation of (1) relationships existing among the components, particularly those of genotypic nature, (2) average gene properties of each of the components and (3) selection indices which may be easily established for any set of the variables.

The chief emphasis of this study lies in the problem of estimating average gene action, and for this reason parents were chosen to give an extremely wide range of expression for all of the traits. With such extremes, it is not only possible to get an over-all genotypic analysis, but also by disregarding extremes, to estimate gene action for any specified range of the character. Wide extremes also demonstrate to full advantage various associations (i.e., genotypic and environmental correlations) which may exist between traits.

Use of non-segregating populations yields an extremely simple, direct, and probably the most accurate method of obtaining estimates of genotypic and environmental correlations and relationships developing from these estimates. The method, as described earlier, involves only components of analyses of variances and covariances of genetically non-segregating lines. Consequently genotypic correlations do not involve the difficult problem of linkage. However, it must be realized also, that if these techniques are to be applied directly to a plant breeding program then all estimates must

be based on the actual breeding material.

With regard to the techniques involved in the problem of estimating gene action, it is of particular interest to contrast results obtained herein, with results obtained by other methods. Tomato fruit size has been studied extensively by numerous workers. Two investigations are of special interest. MacArthur (19), in 1941, constructed an ingenious experiment involving four inbred lines and all possible F_1's in a latin square design. He concluded, "The same genome substitution regularly produces the same phenotypic effect upon fruit size, when that effect is described in geometric terms on a strictly proportional basis, as a percentage increase or decrease."

Powers (23) using Red current (same as our parent 1) and Danmark, as parents and considering the F_1, F_2 and both backcrosses, concluded that " . . . the effects of the genes differentiating weight of fruit are geometrically cumulative". Also through an estimation of genetic variances he suggested the presence of genic dominance. Such results correspond closely with results as found in this study, in which logarithmic gene action was postulated with negative dominance estimated to be of the order $\bar{h} = -.05$ to $-.10$.

Powers has contributed probably more than any other person to genetic elucidation of quantitative inheritance. He has approached the analysis by numerous devices mostly of a statistical nature. In estimating gene action for number of locules, Powers (22) used a linkage test involving genes "grooved" and "oblate". He postulated that gene action was of a geometric nature. The evidence found in this material indicated although there was some evidence for logarithmic gene action, the arithmetic scheme apparently fitted the data better.

As for the remaining characters, no real effort has been made to estimate the type of gene model most closely approximating the data. Powers (24) constructed a P_1-F_1 experiment involving ten inbred lines and examined various traits, some of which were also studied in this thesis (i.e., X_1, X_2, X_3, X_4, X_7, X_9, and X_{10}). However, these results were discussed using arithmetic data only, and it is evident that Powers was not constructing gene models to fit the data, but merely describing the phenotypic F_1 dominance relationships on the basis of arithmetic means. For example, in regard to one of the variables Powers (24) concluded,

> In respect to number of ripe fruit per plant, the range in expression of dominance and heterosis varied from partial dominance of fewer fruits per plant to heterosis for greater number of fruits per plant. This is significant when considering the range in number of ripe fruit per plant varied from 2.9 to 183.

Arithmetic description of the data found in this study is essentially similar, and involves a range of 16 to 1287 ripe fruits per plant. On proceeding to fit a gene model to the data, it was found that the model with logarithmic gene action and considerable positive dominance ($\bar{h} \approx .80$) fits the data well.

While discussing this character it should be pointed out that the P_1-F_1 c.p.r. trend indicated log action immediately on analysis of arithmetic data, whereas with ordinary examination, log action would not be readily apparent, as the association of log action with high degree of dominance has seldom been considered in the past. The main reason, of course, is the difficulty in identifying log action when a high degree of dominance is present.

One of the interesting aspects of fitting gene models to the numerous variables considered in this thesis was that it was not only possible to

assign relatively simple models to the data, but, in general, the models fitted remarkably well. Probably one of the reasons for close correspondence of simple models to average gene action was the use of extended range of character expression. With shorter ranges a different model may be the best. Such was the case with variable X_{10}. This points out the fact that the model specified should be qualified by range and probably also by the material used.

It must be recognized that a model is merely based on an averaging of gene effects and on a character considered alone and as an entity. That is, when a character is broken down into components, the gene action estimation within each component may be quite different from that found in the parent character. The gene action estimated for the parent character is due to the gene action within each of the components together with the morphological interaction which may exist between the parts. Therefore, there is some inconsistency in extrapolating from a descriptive gene model to specifying actual gene action. Thus, the case of over-dominance in yield was shown to be due to the interaction of two characters where partial dominance of one combined with slight negative dominance in the other to give over-dominance in yield. Also, in maturity time an almost perfect case of over-dominance turned out to be a result of the interaction of components in which partial dominance was the rule.

On the other hand, of course, it must be admitted that, for the specific character of maturity time, the description of the average gene action responsible for variation in this character is of an over-dominance nature.

Nevertheless, there is certainly a distinction between the concept

that the heterozygous gene phase produces a greater effect than either homozygous phase, and the concept that the phenotypic effect is due to interaction of different sets of genes or, on a character level, an interaction of different characters. This is not meant to imply that cases of true over-dominance do not exist, as there are a few single gene analyses which support this idea, and, also, some of the various blood antigen alleles would support East's (4) closely related concept of allelic differentiation. However it may be postulated that in complexly inherited quantitative characters, the "mock" dominance effect, as illustrated in yield and maturity time, probably is of considerable importance.

For estimation of average gene action and closely related problems, it is essential to partition variances into components and estimate genotypic relationships. The approach using non-segregating populations, as described in this study, offers a relatively simple, direct, and probably one of the most accurate methods available because of the avoidance of the linkage problem. Other advantages include the practicability of the scheme in which relatively few plants are needed in contrast with the large numbers needed for segregating populations. This is particularly important with plants such as tomatoes which have indeterminant growth. Also many parents may be used, thus broadening the basis for inferences to be derived from the experiments as well as obtaining information on specific parents and also on groups of parents differing in some respects.

CONCLUSIONS

The problem of genetic analysis of complex quantitative characters was attacked first by the resolution of such characters into simpler component parts. X_1 (Yield) of tomatoes was first broken down into, X_2 (number of ripe fruits per plant), and X_3 (average fruit weight). X_2 was then broken down into X_4 (number of clusters bearing ripe fruit per plant) and X_5 (number of fruits per cluster). X_3 was subdivided into X_6 (number of locules per fruit) and X_7 (weight per locule). Three other characters were also included in the study and these were X_8 (flowering date), X_9 (maturity time) and X_{10} (number of flowers per cluster).

A method of analysis involving non-segregating populations, named "constant parent regression method" was discussed from a theoretical point of view and then applied to the problem of estimating the average gene properties of these various characters. Two closely related problems were also approached by use of the statistics derived from non-segregating populations. These problems were (1) estimation of relationships (both genotypic and phenotypic) among the components of yield and (2) establishing a discriminant function selection index for yield components.

A summary of the various gene models is found in Table 36.

Genotypic and phenotypic relationships were established among the components and the genotypic values are summarized in figure 3. Phenotypic relationships were almost identical.

The discriminant function designed to maximize genotypic gain for yield was found to be as follows:

$$Y = .912\,X_4 + .834\,X_5 + .884\,X_6 + .911\,X_7 .$$

Table 36

Summary of best fitting gene models for all characters studied.

Trait	Mode of gene action	Dominance (approximate n values)
X_1 Yield	Arithmetic	+1.00 to+2.41 (complete to over-dominance)
X_2 Number of ripe fruit per plant	Logarithmic	+.80
X_3 Average fruit weight	Logarithmic	-.05 to -.10
X_4 Number of clusters bearing ripe fruit	Logarithmic	+.75
X_5 Number of fruits per cluster	Arithmetic	+.15 to +.20
X_6 Average number of locules per fruit	Arithmetic	-.60
X_7 Weight per locule	Logarithmic	-.05 to -.10
X_8 Flowering date	Arithmetic	-.15
X_9 Maturity time	Arithmetic	-5.24 (over-dominance)
X_{10} Number of flowers per cluster	Logarithmic	+.16

Finally, it must be pointed out that these results were based on material which presented extreme genotypic ranges of expression for all traits, and for most cases, were measured on a logarithmic scale.

It was found that the technique of using non-segregating populations provided an excellent approach to all of these problems, and also had decided theoretical as well as practical advantages over the use of segregating populations.

LITERATURE CITED

1. Charles, D. R., and Smith, H. H. 1939. Distinguishing between two types of gene action in quantitative inheritance. Genetics 24:34-48.

2. Cramer, H. 1946. Mathematical methods of statistics. 1st ed. Princeton University Press, Princeton, New Jersey.

3. Crump, S. L. 1948. The estimation of components of variance in multiple classifications. Unpublished Thesis (Ph.D.), Ames, Iowa. Iowa State College Library.

4. East, E. M. 1936. Heterosis. Genetics 21:376-397.

5. Fisher, R. A., Immer, F. R., and Tedin, O. 1932. The genetical interpretation of statistics of the third degree in the study of quantitative inheritance. Genetics 17:107-124.

6. Frankel, O. H. 1947. The theory of plant breeding for yield. Heredity 1:109-120.

7. Gowen, J. W. 1947. Sex determination in Drosophila melanogaster. Records of Genet. Soc. Amer. 16:33.

8. Griffing, J. B. 1947. Statistical analysis of components of genetic variance in maize in relation to heterosis. Unpublished Thesis (M.S.,) Ames, Iowa. Iowa State College Library.

9. Harland, S. C. 1919. The improvement of the yield of Sea Island cotton in the West Indies by the isolation of pure strains. Parts 1 and 2. West Indian Bul. 17:145-161, 210-236.

10. _____ 1939. Genetical studies in the genus Gossypium and their relationship to evolutionary and taxonomic problems. Proc. 7th Int. Genet. Congr. 7:138-143.

11. _____ 1944. The selection experiment with Peruvian tanguis cotton. Sociedad Nacional Agraria, Institute of Cotton Genetics, Lima, Peru. Bul. 1.

12. Hazel, L. N. 1943. Genetic basis for constructing selection indices. Genetics 28:476-490.

13. _____, and Baker, M. L. and Reinmiller, C. F. 1943. Genetic and environmental correlations between the growth rates of pigs at different ages. Jour. Anim. Sci. 2:118-128.

14. Hull, F. 1945. Lectures presented at genetics seminars. Iowa State College, Ames, Iowa. (Unpublished)

15. Khambanonda, I. 1948. Polygenic inheritance of fruit size in red pepper (Capsicum fruitescens L.) Unpublished Thesis (Ph.D.) Ames, Iowa. Iowa State College Library.

16. Lindstrom, E. W. 1926. Hereditary correlation of size and color characters in tomatoes. Ia. Agr. Exp. Sta. Bul. 93.

17. Lush, J. L. 1945. Animal breeding plans. 3rd ed. Ames, Iowa, Collegiate Press, Inc.

18. _____ 1948. The genetics of populations. Iowa State College Ames, Iowa. (Mimeographed)

19. MacArthur, J. W. 1941. Size inheritance in tomato fruits. Jour. Hered. 32:291-295.

20. McCleskey, C. S. 1926. Biometrical studies on quantitative characters in tomatoes. Unpublished Thesis (M.S.), Ames, Iowa. Iowa State College Library.

21. Panse, V. G. 1940. Application of genetics to plant breeding. II. The inheritance of quantitative characters and plant breeding. Jour. Genet. 40:283-302.

22. Powers, L. 1939. Studies on the nature of the interactions of the genes differentiating quantitative characters in a cross between Lycopersicon esculentum and L. pimpinellifolium. Jour. Genet. 39:139-170.

23. _____ 1942. The nature of the series of environmental variances and the estimation of the genetic variances and the geometric means in crosses involving species of Lycopersicon. Genetics 27:561-575.

24. _____ 1945. Relative yields of inbred lines and F_1 hybrids of tomato. Bot. Gaz. 106:247-268.

25. Rasmusson, J. 1933. A contribution to the theory of quantitative character inheritance. Hereditas 18:245-261.

26. Richey, F. D. 1942. Mock-dominance and hybrid vigor. Science 96:280-281.

27. Smith, H. F. 1936. A discriminant function for plant selection. Ann. Eugenics 7:240-250.

28. Snedecor, G. W. 1946. Statistical methods. 4th ed. Iowa State College Press, Ames, Iowa.

29. Stern, C. and Schaeffer, E. W. 1943. On wild type iso-alleles in Drosophila melanogaster. Proc. Nat. Acad. Sci. 29:361-367.

30. Weber, C. R. 1948. Inheritance and interrelation of some agronomic and chemical characters in an inter-specific cross in soybeans, Glycine max x G ussurienses. Unpublished Thesis (Ph.D.), Ames, Iowa. Iowa State College Library.

31. Wright, S. 1934. The method of path coefficients. Ann. Math. Stat. 5:161-215.

32. Yeager, A. F. 1936. Genetical and histological studies of fruit size and shape in the tomato. Unpublished Thesis (Ph.D.), Ames, Iowa. Iowa State College Library.

ACKNOWLEDGMENTS

This study was initiated, developed, and experimentally completed under the supervision of Dr. E. W. Lindstrom, former head, of the Genetics Department, Iowa State College. His death occurred shortly before the thesis material was assembled in final form. To acknowledge the indebtedness of the author to Dr. Lindstrom in a few sentences is impossible. His guidance and counsel embraced not only this thesis problem but also the entire undergraduate and graduate educational training of the author.

The writer is indebted to Dr. Gowen and Dr. Lush for valuable criticisms in regard to the text. The author is particularly grateful, also, to Professor O. Kempthorne for criticisms of the statistical aspects of the thesis. Acknowledgment is also given to Professor Kempthorne and Dr. A. Robertson for the suggestion of the form of prediction equation (1) on page 34.

APPENDIX

Table 37

Individual plant values for X_1 (Yield) for parents, F_1, F_2, B_1, and B_2 involving cross Devon x Matchless. Values coded and data adjusted for replication effects.

Generation	Total plant yield in grams. (coded)																													Mean*
	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	
(5)Devon						1	-	4	1	3	2	5	1	-	1																1996
(6)Matchless								1	1	1	2	1	2	3	2	1	2	1													2253
(5x6)F_1											2	3	2	3	2	1	-	1	2	-	2										2452
F_2	2	1	2	1	2	5	10	2	9	10	14	16	12	12	4	9	5	1	2	2	-	1	-	1	-	1	-	-	-	1	2135
(5x6)x6				2	2	-	1	2	2	5	5	10	9	5	6	8	8	7	-	4	3	2	-	-	1	-	1	-	1		2375
(5x6)x5	2	-	3	-	2	6	9	3	5	7	9	2	10	2	3	2	6	1	2	2	4	-	1	-	-	1	1				2121

* not coded

Table 38

Individual plant values for X_2(Number of fruits per plant) of parents, F_1, F_2, B_1, and B_2 for cross Devon x Matchless. Both unadjusted and adjusted values for riplication effects given.

Generation		Number of fruits per plant																							Mean
		10	13	16	19	22	25	28	31	34	37	40	43	46	49	52	55	58	61	64	67	70	73		
5	adj.									1	3	5	8	1											
	unadj.					1	-	1	1	1	1	6	-	3	2	2									40.6
6	adj		4	11	3																				
	unadj.	2	4	6	5	1																			15.9
5x6	adj							1	3	6	1	4	2	1											
	unadj.							4	1	4	2	2	3	2											36.3
F_2	adj.			3	5	11	15	11	15	19	11	16	7	3	4	1	2	3							
	unadj.		1	2	7	11	8	19	18	14	13	13	3	7	2	4	-	-	2	1					32.9
(5x6)x5	adj.					1	2	5	7	13	8	15	8	6	8	4	5	1	-	1					
	unadj.			1	-	-	5	5	7	12	11	13	6	5	5	3	2	5	-	2	1	-	1		40.4
(5x6)x6	adj.	2	5	6	12	10	21	11	11	2	1	1	-	-	1										
	unadj.	6	4	4	17	9	20	13	3	3	3	-	-	1											23.2

Table 39

Individual plant values for X_3 (Average fruit weight) of parents, F_1, F_2, B, and B_2 for the cross Devon x Matchless. Values adjusted for replication effects.

Generation	Fruit weight in grams. 30	35	40	45	50	55	60	65	70	75	80	85	90	95	100	5	10	5	20	5	30	5	40	5	50	5	60	5	Mean
5			1	11	5	1																							49.1
6																		1	-	2	2	3	3	3	1	1	1	1	142.6
5x6						1	4	9	3	1																			67.3
F_2	1	1	7	8	14	8	12	20	18	7	3	6	2	3	2	1	2												67.8
(5x6)x5				4	16	25	14	8	11	5	1																		58.7
(5x6)x6						1	2	2	6	3	9	12	7	7	10	6	6	6	1	3	-	1	-	-	-	1			93.5

Table 40

Individual plant values for X_4 (Number of clusters) for parents, F_1, F_2, B_1, and B_2 involving cross Devon x Matchless. Both unadjusted and adjusted values for replication effects given.

Generation	Number of clusters bearing ripe fruit per plant																															Means
	2.2	2.8	3.4	4.0	4.6	5.2	5.8	6.4	7.0	7.6	8.2	8.8	9.4	10.0	10.6	11.2	11.8	12.4	13.0	13.6	14.2	14.8	15.4	16.0	16.6	17.2	17.8	18.4				
5 adj.						1	-	-	-	2	3	1	7	2	-	1	-	1														
unadj.								2	-	3	4	3	3	-	-	1	-	1	-	-	-	-	1									9.0
6 adj.				1	1	5	4	1	4	-	1	-	1																			
unadj.				1	2	4	5	-	3	1	1	-	-	1																		6.1
5x6																																
adj.									3	-	2	1	1	5	1	2	1	1	1													
unadj.						1	1	1	1	2	1	2	2	1	-	1	1	1	1	-	-	1	-	-	-	1						9.7
F_2 adj.				1	3	4	10	8	10	10	16	6	20	8	10	5	3	5	1	1	1	1	1	-	1							
unadj.				2	3	6	6	4	11	23	10	19	8	8	1	5	7	2	3	3	2	-	-	-	1	-	-	1				8.7
(5x6)x5																																
adj.					2	1	5	5	8	8	11	11	5	9	5	7	2	-	1	1	2	-	-	1								
unadj.					2	4	5	11	9	8	4	7	5	5	6	3	4	2	3	2	-	1	1	2								8.8
(5x6)x6																																
adj.	1	2	-	1	1	1	10	13	13	9	12	6	4	-	4	3	1	1	-	-	-	1										
unadj.				4	3	7	14	6	11	7	6	5	5	5	3	-	2	2	1	-	1	-	-	-	1							7.5

Table 41

Individual plant values for X_5 (Number of fruits per cluster) for parents, F_1, F_2, B_1 and B_2 involving cross Devon x Matchless. Both unadjusted and adjusted values for replication effects given.

Generation	Number of fruits per cluster																														Means
	1.0	1.3	1.4	1.9	2.1	2.5	2.8	3.1	3.4	3.7	4.0	4.3	4.6	4.9	5.2	5.5	5.8	6.1	6.4	6.7	7.0	7.3	7.6	7.9	8.2	8.5	8.8	9.1	9.4	9.7	
5 adj.							1	-	2	3	4	1	6	5	3	2	3	4	3	1	2	1	-	-	-	-	-	1			
unadj.						2	1	1	2	1	5	1	2	6	1	5	1	5	5	-	1	-	2	-	-	-	-	-	-	-	1	5.2
6 adj.	2	-	4	4	4	6	8	8	3	1	2																				
unadj.	2	-	3	9	-	9	-	11	5	-	3																				2.6
5x6 adj.						1	1	1	7	2	8	5	3	10	2	1	-	1													
unadj.				1	-	1	2	2	9	2	6	1	3	6	-	8	-	1													4.2
F_2 adj.				2	8	5	4	4	12	22	16	19	9	6	7	4	6	-	-	-	-	1									
unadj.			4	4	1	5	1	12	19	7	26	3	17	12	-	7	-	4	4												3.9
(5x6)x5 adj.						1	2	4	6	4	2	8	15	7	10	6	8	5	2	1	2										
unadj.						1	1	4	8	2	8	2	11	21	2	9	-	10	3	-	2										4.8
(5x6)x6 adj.		1	3	2	6	-	19	16	7	15	4	8	1	1	1																
unadj.	2	-	4	11	-	9	-	10	22	-	10	-	11	4																	3.2

Table 42

Individual plant values for X_6 (Number of locules per fruit) for parents F_1, F_2, B_1 and B_2 involving cross Devon x Marchless. Unadjusted data only.

Generation	Number of locules per fruit																															Means
	2.05	.25	.45	.65	.85	3.05	.25	.45	.65	.85	4.05	.25	.45	.65	.85	5.05	.25	.45	.65	.85	6.05	.25	.45	.65	.85	7.05	.25	.45	.65	8.25	
5		1	8	22	6	4	1																									2.7
6																				2	1	2	8	7	8	4	5	3	1	1		6.6
5x6						1	7	9	9	8	8																					3.7
F_2		1	7	10	9	4	10	6	17	5	10	10	7	7	6	4	3	4	3	1	1											3.9
(5x6)x5	1	3	9	8	8	7	12	14	13	7	1	-	-	1																		3.2
(5x6)x6						1	5	1	7	8	4	6	-	2	5	3	6	6	5	8	3	3	4	4	2							5.0
Total	1	5	24	40	23	17	35	30	46	28	23	16	7	10	11	7	9	10	8	11	5	5	12	11	10	4	5	3	1	1		

Table 43

Individual plant values for X_7 (Weight per locule) for parents, F_1, F_2, B_1 and B_2 involving cross Devon x Matahless. Adjusted values for replication effects only.

Generation	Weight in grams per locule																														Means
	12.5	13.0	13.5	14.0	14.5	15.0	15.5	16.0	16.5	17.0	17.5	18.0	18.5	19.0	19.5	20.0	20.5	21.0	21.5	22.0	22.5	23.0	23.5	24.0	24.5	25.0	25.5	26.0	26.5	27.0	27.5	28.0	
5							1	1	1	2	2	3	3	1	1	1	1	1															18.1
6										1	-	-	2	2	2	2	1	1	3	1	1	1	-	-	1								20.5
5x6				1	-	-	2	-	1	-	-	4	3	-	5	1	-	1															18.2
F_2	1	1	3	3	4	6	10	13	9	10	12	8	7	8	3	8	-	4	2	1	4	3	2	-	-	-	1	-	2				17.8
(5x6)x5		1	-	1	2	3	-	9	5	8	3	4	6	3	9	7	6	5	3	2	2	3	-	1	-	-	-	-	-	-	-	1	18.8
(5x6)x6				1	3	4	1	4	4	8	5	5	4	5	3	8	4	5	5	2	5	3	-	-	-	1	2	1					19.1

Table 44

Individual plant values for X_8 (Flowering date) for parents F_1, F_2, B_1 and B_2 involving cross Devon x Matchless. Unadjusted and adjusted values for replication effects given.

Generation	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	Means
5 adj.							1	-	-	7	8	8	9	5	1	2												16.1
unadj.						1	-	2	6	2	4	6	8	4	6	1	1											
6 adj.								1	-	2	6	8	8	13	2	-	-	-	1									
unadj.								1	-	3	9	6	8	9	3	1	-	-	1									16.8
5x6 adj.								1	2	8	11	7	5	4	1	3												
unadj.							1	2	5	2	11	10	4	1	4	1	1											15.6
F_2 adj.						2	2	2	2	11	11	8	21	20	16	11	6	7	2	1	1	-	1					
unadj.		1	-	-	-	4	1	6	4	10	21	20	18	11	5	7	5	5	2	2	-	-	-	1	-	1		17.6
(5x6)x5																												
adj.	1	-	1	-	1	-	2	2	12	4	17	9	19	7	5	2												
unadj.			1	1	1	1	1	9	5	21	13	16	8	1	4													15.4
(5x6)x6																												
adj.								2	8	2	7	14	11	10	14	9	5											
unadj.									9	3	7	13	11	6	17	10	5	1										17.3

Table 45

Individual plant values for X_9 (Maturity time) for parents, F_1, F_2, B_1 and B_2 involving cross Devon x Matchless. Unadjusted and adjusted values for replication effects given.

Generation	Number of days from first flower to first fruit ripe																						Means
	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	
5 adj.						1	1	2	9	12	8	4	4	-	-	-	-	1						
unadj.					1	1	1	-	8	9	12	3	6	-	-	-	1							42.6
5 adj.								1	1	3	5	10	10	5	5	-	2							
unadj.								1	4	5	3	9	3	6	1	6	3	1						44.8
5x6 adj.				1	-	-	-	4	7	13	6	3	4	1	1									
unadj.				1	-	-	-	4	7	13	6	3	4	1	1									42.3
F_2 adj.	1	-	-	-	-	-	3	8	17	26	20	19	14	6	7	2	-	-	-	1				
unadj.	1	-	-	-	-	1	4	7	16	23	24	18	16	2	8	3	-	-	-	-	1			43.1
(5x6)x5 adj.						2	3	11	12	18	17	11	6	2	1									
unadj.						1	7	8	15	14	17	13	3	4	-	1								42.2
(5x6)x6 adj.							3	3	9	7	15	16	13	5	4	2	2	-	-	2				
unadj.						1	2	5	3	15	10	17	12	6	4	2	-	2	-	1	1			43.9

Table 46

Individual plant values for X_{10} (Number of flowers per cluster) for parents, F_1, F_2, B_1 and B_2 involving cross Devon x Matchless. Unadjusted and adjusted values for replication effects given.

Generation	Number of flowers per cluster: 3.4	3.7	4.0	4.3	4.6	4.9	5.2	5.5	5.8	6.1	6.4	6.7	7.0	7.3	7.6	7.9	8.2	8.5	8.8	9.1	9.4	9.7	10.0	10.3	10.6	10.9	Mean
5 adj.						1	-	-	2	-	4	5	3	4	8	6	6	2	-	-	-	-	-	1			
unadj.						1	1	-	-	2	2	4	8	4	7	7	3	-	1	1	-	-	-	-	-	1	7.4
6 adj.	2	1	6	17	7	8	-	1																			
unadj.	1	3	8	13	8	6	2	-	1																		4.4
5x6 adj.						1	2	5	13	11	8	1	-	1													
unadj.						2	4	1	9	12	10	3	-	1													6.0
F_2 adj.			1	5	3	16	16	20	16	20	13	4	9	1	-	-	1										
unadj.				7	12	18	18	1	17	16	18	10	6	-	1	1											5.7
(5x6)x5 adj.						4	3	11	8	15	7	5	11	7	3	8	1	1									
unadj.					3	2	9	0	6	14	13	11	7	5	10	2	1	-	-	1							6.5
(5x6)x6 adj.	1	1	9	9	10	26	12	10	4	1																	
unadj.	2	2	7	8	25	14	14	-	7	4																	4.8

Table 47

Arithmetic means for parents and F_1's for all ten characteristics studied.

P_1 or F_1 Number	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
1	678.	1287.	.5	133.5	10.5	2.00	.3	5.0	38.3	19.6
2	921.	358.	2.6	66.8	5.6	2.01	1.3	5.5	40.2	7.5
3	1144.	260.	4.4	41.7	6.2	2.00	2.2	11.7	41.1	7.2
4	1628.	50.	32.7	13.4	3.9	2.09	15.4	20.4	40.7	4.9
5	1996.	41.	49.1	8.9	5.2	2.68	18.2	16.2	42.6	7.4
6	2253.	16.	142.6	6.1	2.3	6.81	20.5	16.8	44.9	4.4
1x2	1198.	1187.	1.0	134.4	9.4	2.00	.5	4.6	39.1	10.6
1x3	1681.	1201.	1.4	129.1	9.7	2.00	.7	6.8	37.5	11.3
1x4	2523.	661.	3.8	77.5	8.7	2.00	1.9	10.7	34.1	10.1
1x5	2552.	612.	4.2	93.5	8.4	2.00	2.1	8.4	35.5	11.3
1x6	2456.	561.	4.4	81.7	7.6	2.13	2.1	8.1	35.4	9.5
2x3	1437.	442.	3.2	84.0	6.0	2.01	1.6	9.8	38.1	7.4
2x4	1824.	226.	8.1	45.1	5.0	2.03	4.0	13.5	33.8	6.3
2x5	1953.	188.	10.3	39.7	5.6	2.11	4.9	11.5	35.4	7.1
2x6	2320.	155.	14.9	35.6	4.7	3.17	4.8	11.6	35.3	6.2
3x4	2539.	212.	12.0	37.3	5.7	2.03	5.9	14.7	37.0	6.5
3x5	2409.	193.	12.5	32.0	6.2	2.16	5.8	14.0	39.1	7.3
3x6	2265.	139.	16.3	28.1	5.2	3.23	5.0	14.9	38.3	6.1
4x5	2683.	60.	45.0	13.0	4.9	2.41	19.0	16.4	40.5	6.5
4x6	2341.	32.	73.2	8.7	3.5	3.83	19.0	18.4	42.9	4.9
6x8	2452.	36.	67.3	9.7	4.2	3.65	18.2	15.5	42.5	6.0